SPE 1/RM/46 – Mars 2005
Section de l'élaboration et de l'application des méthodes
Centre de technologie environnementale
Environnement Canada

SÉRIE DE LA PROTECTION DE L'ENVIRONNEMENT

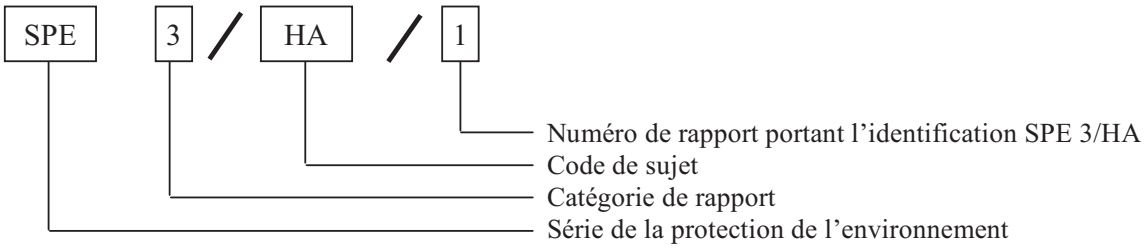
Document d'orientation sur les méthodes statistiques applicables aux essais d'écotoxicité

Environnement Canada Environment Canada

Canada

SÉRIE DE LA PROTECTION DE L'ENVIRONNEMENT

Exemple de numérotation



Catégories

1 Règlements/Lignes directrices/Codes de pratiques
2 Évaluation des problèmes et options de contrôle
3 Recherche et développement technologique
4 Revues de la documentation
5 Inventaires, examens et enquêtes
6 Évaluations des impacts sociaux, économiques et environnementaux
7 Surveillance
8 Propositions, analyses et énoncés de principes généraux
9 Guides

Sujets

AG Agriculture
AN Technologie anaérobie
AP Pollution atmosphérique
AT Toxicité aquatique
CC Produits chimiques commerciaux
CE Consommateurs et environnement
CI Industries chimiques
FA Activités fédérales
FP Traitement des aliments
HA Déchets dangereux
IC Produits chimiques inorganiques
MA Pollution marine
MM Exploitation minière et traitement des minéraux
NR Régions nordiques et rurales
PF Papier et fibres
PG Production d'électricité
PN Pétrole et gaz naturel
RA Réfrigération et conditionnement d'air
RM Méthodes de référence
SF Traitement des surfaces
SP Déversements de pétrole et de produits chimiques
SRM Méthodes de référence normalisées
TS Transports
TX Textiles
UP Pollution urbaine
WP Protection et préservation du bois

Des sujets et des codes additionnels sont ajoutés au besoin. On peut obtenir une liste des publications de la Série de la protection de l'environnement à l'adresse suivante : Services des communications, Environnement Canada, Ottawa (Ontario) K1A 0H3.

Document d'orientation sur les méthodes statistiques applicables aux essais d'écotoxicité

Section de l'élaboration et de l'application des méthodes
Centre de technologie environnementale
Environnement Canada
Ottawa

Rapport SPE 1/RM/46
Mars 2005

Catalogage avant publication (sous la coordination de Bibliothèque et Archives Canada)

Document d'orientation sur les méthodes statistiques applicables aux essais d'écotoxicité /
Section de l'élaboration et de l'application des méthodes, Centre de technologie environnementale,
Environnement Canada

(Rapport ; SPE 1/RM/46)
Comprend un résumé en anglais.
Publié aussi en anglais sous le titre : Guidance Document on Statistical Methods for Environmental Toxicity Tests
Également disponible sur l'Internet.
Méthode d'essai biologique. Cf. l'avant-propos.
Comprend des références bibliographiques : 304 p.
ISBN 0-660-97065-1
N° de cat. : En49-7/1-46F

1. Toxicité — Méthodes statistiques.
2. Écotoxicologie — Méthodes statistiques.
3. Toxicologie expérimentale — Méthodes statistiques.
4. Essais biologiques — Méthodes statistiques.
5. Eau — Qualité — Essais biologiques — Méthodes statistiques.
6. Toxicité — Canada — Méthodes statistiques.
I. Canada. Environnement Canada
II. Centre de technologie environnementale (Canada). Section de l'élaboration et de l'application des méthodes
III. Titre : Méthode d'essai biologique.
IV. Coll. : Rapport (Canada. Environnement Canada) SPE 1/RM/46.

QK46.5 S7 B5614 2005 615.9'02'0727 C2005-980198-0

© Sa Majesté du chef du Canada (Environnement Canada) 2005
N° de catalogue En49-7/1-46F
ISBN 0-660-97065-1

Commentaires

Adresser les commentaires sur la teneur du présent rapport à :

Richard P. Scroggins
Chef, Division des méthodes biologiques
Centre de technologie environnementale
Environnement Canada
335, River Road
Ottawa
K1A 0H3

This report is also available in English from:

Environmental Protection Publications
Environment Canada
Ottawa (Ontario)
K1A 0H3

Avis de révision

Le présent document a été révisé par le personnel de la Direction générale de l'avancement des technologies environnementales d'Environnement Canada, et sa publication a été autorisée. La mention d'appellations commerciales ou de produits offerts sur le marché ne constitue ni une approbation de ces produits ni une recommandation de leur emploi par Environnement Canada. D'autres produits de valeur semblable existent.

Résumé

Le présent document d'orientation étaye et complète les méthodes applicables aux essais de toxicité monospécifiques publiées par Environnement Canada. Il s'adresse en particulier au nouveau personnel de laboratoire.

Ce document donne des conseils supplémentaires sur l'analyse statistique des résultats des essais d'Environnement Canada. On y trouvera des observations sur les modes opératoires souhaitables et les pièges courants. Il rappelle des notions de statistique, mais ce n'est pas une initiation à la statistique. Il ne tente pas non plus d'innover dans le domaine de l'analyse statistique, bien qu'il attire l'attention sur des méthodes qui sont en cours de développement et qui semblent prometteuses. Il aborde les méthodes applicables aux essais de toxicité létale et sublétale, en insistant davantage sur les essais, plus nombreux, en milieu aquatique (colonne d'eau et sédiments).

Outre un glossaire détaillé, le document renferme une section sur les plans d'expérience. Cette section souligne la nécessité de consulter un statisticien, de choisir les concentrations, de toujours utiliser le logarithme de la concentration, il insiste sur les divers types de témoins, les toxiques de référence, la randomisation, les répétitions ainsi que la transformation des données.

Le document expose les essais à concentration unique parmi lesquels on peut choisir, de même que les limites imposées par le plan d'expérience.

Une section sur les essais de toxicité quantique décrit les méthodes d'estimation des concentrations efficaces et de leurs limites de confiance ainsi que la conduite à tenir à l'égard des effets observés chez les témoins. Avec de bonnes données, diverses méthodes d'analyse arrivent à des paramètres de toxicité semblables. On recommande la régression probit, s'il y a deux effets partiels, de préférence à l'aide de techniques du maximum de vraisemblance. Le choix s'arrête sur la méthode de Spearman-Kärber avec équeutage limité, s'il n'y a qu'un effet partiel, et sur la méthode binomiale, si aucun effet n'est partiel et que les effets sont nuls ou totaux. Pour déceler les erreurs, on devrait tracer la droite à la main. Les courbes de toxicité et les analyses des temps efficaces présentent des avantages.

Pour les essais quantitatifs, qui portent habituellement sur un effet sublétal, la méthode préférée est une estimation ponctuelle, par régression, de la concentration inhibitrice (CI). Environnement Canada a récemment exigé comme premier choix pour l'analyse la régression linéaire et non linéaire (v. le § 6.5.8). Cette analyse remplace l'estimation de la CI par lissage et interpolation (programme ICPIN) qui était couramment utilisée. Le test d'hypothèse visant à déterminer une concentration « sans effet observé » est exposé en détail en raison de son utilisation si fréquente. Cette méthode est beaucoup moins souhaitable que les estimations ponctuelles, et son utilisation est en recul.

Dans les essais de mesure d'un double effet, la corrélation entre les deux effets et leurs différentes distributions statistiques créent de graves problèmes d'analyse. L'approche la plus opportune consiste à séparer l'analyse de l'effet quantitatif (habituellement sublétal) de celle de l'effet quantique (habituellement létal). Une autre façon, que l'on peut justifier écologiquement, consiste à combiner les deux effets dans une analyse de la biomasse. D'habitude, cela donne un effet plus marqué.

Les notions de statistique comprennent une discussion des difficultés engendrées par l'habituelle estimation « inversée » des paramètres de toxicité et de leurs limites de confiance. On décrit des méthodes restreintes permettant de tester les différences significatives entre deux et plusieurs paramètres de toxicité et la conduite à tenir à l'égard des observations aberrantes. On donne des conseils sur l'interprétation d'autres relations dose-effet aberrantes.

Abstract

This guidance document supports and supplements the methods for single-species toxicity tests, published by Environment Canada. In particular, it is intended for new laboratory personnel.

This document provides additional guidance for statistical analysis of results from Environment Canada tests. It comments on desirable procedures and common pitfalls. Some statistical background is covered, but this document does not teach basic statistics. Nor does it attempt to break new ground in statistical analysis, although it points to methods that are under development and seem promising for future use. This document covers methods for lethal and sublethal tests, with most emphasis on the more numerous aquatic tests (water-column and sediment).

A detailed glossary is provided. A design chapter emphasizes the need for consultation with a statistician, choice of concentrations, staying with logarithm of dose, the various types of controls, reference toxicants, randomization, replication, and transformation of data.

Choices among single-concentration tests are outlined, and the limitations imposed by design.

A section on quantal tests outlines methods for estimating effective concentrations (ECp) and confidence limits, and dealing with control effects. Various analytical methods provide similar endpoints for good data. Probit regression is recommended if there are two partial effects, preferably by maximum likelihood techniques. The Spearman-Kärber method with limited trimming is the choice if there is only one partial effect, and the binomial method if only zero and complete effects are available. A line should be plotted by hand to check for errors. Toxicity curves and analyses of effective times are beneficial.

For quantitative tests, which are usually sublethal, a point-estimate of the inhibition concentration (ICp) by regression is the most favoured method. Environment Canada has recently required linear and nonlinear regression as the first choice for analysis (Section 6.5.8). That analysis replaces the estimation of ICp by smoothing and interpolation (the ICPIN program) which has been commonly used. Hypothesis testing to determine a "no-observed-effect" concentration (NOEC) is outlined in detail because it has been used so frequently; this approach is much less desirable than point-estimates, and its use is decreasing.

In dual-effect tests, the correlation between the two effects, and their different statistical distributions, creates severe analytical problems. The most expedient approach is to separate the analysis of the quantitative component (usually sublethal) from the analysis of the quantal effect (usually lethal). An alternative approach that can sometimes be justified on ecological grounds is to combine the two effects into a "biomass" analysis, an approach that usually produces a more pronounced effect.

The statistical background includes discussion of difficulties caused by the customary "inverse" estimation of endpoints and confidence limits. Limited methods are described for testing significant differences between and among endpoints, and dealing with outliers. Advice is given for interpreting other deviant dose-effect relationships.

Avant-propos

*Le présent document est publié dans la collection des **guides** d'Environnement Canada (EC) qui portent sur les méthodes recommandées ou normalisées d'essai biologique. Pour les essais dont il est question, on utilise une seule espèce aquatique ou terrestre et on se place dans les conditions contrôlées du laboratoire pour mesurer les effets toxiques d'échantillons de matières choisies. Les méthodes recommandées ont été évaluées par Environnement Canada et elles sont privilégiées dans les cas suivants :*

- *pour être utilisées dans les laboratoires d'Environnement Canada dans des essais d'écotoxicité ;*
- *pour les essais impartis par Environnement Canada ou demandés par des organismes de l'extérieur ou l'industrie ;*
- *pour inspirer des directives très explicites qui pourrait être formulées dans une méthode de référence ou une méthode réglementaire normalisée.*

Les différents types d'essais traités dans la collection se sont révélés convenir aux besoins des programmes de gestion et de protection de l'environnement exécutés par Environnement Canada. Les descriptions des méthodes d'essai visent à orienter et à faciliter l'emploi de modes opératoires cohérents, appropriés et complets pour l'obtention de données sur la toxicité pour les organismes aquatiques et terrestres. Les essais visent à permettre l'évaluation de matières simples ou complexes, destinées à être rejetées dans l'environnement ou qui sont déjà présentes dans un milieu donné tel que les sédiments.

*Dans l'annexe A, on énumère les **méthodes d'essai biologique** génériques (universelles), polyvalentes, les **méthodes de référence** normalisées et les **guides** à l'appui, qui ont été publiés jusqu'à présent. Ces méthodes et guides, produits par la Section de l'élaboration et de l'application des méthodes d'Environnement Canada, à Ottawa, peuvent être obtenues de Publications de la Protection de l'environnement, Environnement Canada, Ottawa, K1A 0H3, Canada. Les conseils figurant dans les documents sont partagés et appliqués par conjointement par les bureaux régionaux et l'administration centrale d'Environnement Canada, dont les coordonnées se trouvent dans l'annexe C.*

Table des matières

Résumé v
Abstract vi
Avant-propos vii
Liste des tableaux xiv
Liste des figures xiv
Abréviations et symboles xvi
Glossaire xvii
Remerciements xli

Section 1
Introduction 1
1.1 Buts et objectifs du document 1
1.2 Mode d'emploi du document 2
1.3 Principales catégories d'essais 3

Section 2
Planification générale et analyse 6
2.1 Participation d'un statisticien 6
2.2 Sélection des concentrations 6
2.2.1 Influences contraires 7
2.2.2 Types particuliers d'essais 8
2.3 Logarithmes de la concentration 9
2.3.1 Constance dans l'emploi des logarithmes 10
2.3.2 Logarithmes et programmes informatiques 11
2.3.3 Calculs ultérieurs avec des logarithmes 12
2.3.4 Cela importe-t-il ? 12
2.3.5 Familiarisation et techniques 13
2.3.6 Logarithme du temps 14
2.3.7 Logarithme de l'effet ? 14
2.4 Randomisation 15
2.5 Répétitions et nombre d'organismes 17
2.5.1 Terminologie 17
2.5.2 Répétition dans les diverses sortes d'essais 18
2.5.3 Relations avec l'échantillonnage sur le terrain 20
2.6 Pondération 21
2.7 Témoins 22
2.7.1 Témoins ordinaires 23
2.7.2 Témoins du solvant 23
2.7.3 Témoins de la salinité 24
2.7.4 Sédiments et sols témoins et de référence 26
2.8 Toxiques de référence et cartes de contrôle 26
2.8.1 Variation raisonnable 27
2.9 Transformation des données sur l'effet 29
2.9.1 Utilisation en régression 30
2.9.2 Utilisation pour le test d'hypothèse 31
2.9.3 Transformations particulières 32

Section 3
Essais à concentration unique 33

3.1 Effets quantiques 34
3.1.1 Un seul échantillon sans répétition 34
3.1.2 Répétition au même emplacement 36
3.1.3 Lieux de prélèvement d'échantillons multiples 36
3.2 Effets quantitatifs à un endroit 37
3.3 Essais quantitatifs sur échantillons provenant de plusieurs endroits 38
3.3.1 Tests paramétriques 39
3.3.2 Tests non paramétriques 39

Section 4
Essais quantiques pour estimer la CE p 41
4.1 Les paramètres de toxicité estimés au moyen d'essais quantiques 43
4.2 Marche à suivre pour toutes les méthodes d'estimation d'une CE p 44
4.2.2 Transformation log-probit 46
4.2.3 Estimation de la CE 50 à l'aide d'un graphique tracé à la main 46
4.2.4 Effets chez les organismes témoins 49
4.2.5 Limites de confiance de la CE p 54
4.2.6 CE 20 ou autres concentrations que la CE 50 57
4.3 Choix de méthodes 58
4.4 Comparaison des estimations par diverses méthodes 60
4.4.1 Estimations faites à l'aide de « bonnes » données 61
4.4.2 Estimations avec des données comportant peu d'effets partiels 65
4.5 Examen des méthodes statistiques d'estimation des CE p 66
4.5.1 Régressions probit et logit en général 69
4.5.2 Autres transformations 70
4.5.3 Régression probit classique informatisée 70
4.5.4 Évaluation de l'ajustement avec le khi-deux 72
4.5.5 Estimations du maximum de vraisemblance 72
4.5.6 Méthode de Spearman-Kärber 73
4.5.7 Méthode binomiale 75
4.5.8 Méthode graphique de Litchfield-Wilcoxon 76
4.5.9 Interpolation linéaire 77
4.5.10 Méthode de la moyenne mobile 77
4.6 Évaluation de nouveaux programmes informatiques 78
4.7 Méthodes non linéaires et autres méthodes possibles de l'avenir 79

Section 5
Temps efficaces, courbes de toxicité et analyse de la survie 80
5.1 Temps efficaces 50 80
5.2 Courbes de toxicité et seuils d'effet 83
5.3 Modélisation des temps efficaces et courbes de toxicité 87
5.4 Analyses de la survie au fil du temps 88
5.4.1 Taux de mortalité 88
5.4.2 Analyse de la survie 88
5.4.3 Mesures répétées 89

Section 6
Estimations ponctuelles pour les essais quantitatifs de toxicité sublétale 90
6.1 Généralités sur les essais de toxicité sublétale 90
6.1.1 Types de tests et de paramètres de toxicité 90
6.2 Rudiments des estimations ponctuelles de paramètres de toxicité sublétale 93
6.2.1 Terminologie 95

6.2.2 Avantages des estimations ponctuelles 95
6.2.3 Répétitions 96
6.2.4 Choix du degré d'effet pour le paramètre de toxicité 96
6.2.5 Sélection de la variable biologique comme paramètre de toxicité 97
6.3 Étapes générales de l'estimation d'un paramètre de toxicité sublétale 98
6.3.1 Tracé des données 98
6.3.2 Choix de la méthode 98
6.4 Lissage et interpolation 98
6.4.1 Critique générale 99
6.4.2 Étapes de l'analyse 100
6.4.3 Le programme informatique ICPIN 100
6.5 Estimations ponctuelles par régression 101
6.5.1 Le b.a.-ba de la régression 102
6.5.2 Notions sur les modèles linéaires, non linéaires, linéaires généraux (GLM) et linéaires généralisés (GLIM) 102
6.5.3 Régression linéaire 103
6.5.4 Aspects généraux des régressions non linéaires 104
6.5.5 Choix d'un modèle de régression 106
6.5.6 Adéquation et ajustement 107
6.5.7 Exemples récents de régressions non linéaires 108
6.5.8 La méthode de régression d'Environnement Canada 109
6.5.9 Un nouveau programme de régression : Newtox-Logstat 114
6.5.10 Modèles linéaires généraux 115
6.5.11 Modèles linéaires généralisés 116
6.5.12 Reparamétrage 117
6.5.13 Autres exemples de tentatives de régression 118
6.6 Seuils estimés par régression 118
6.6.1 Seuils estimés par le modèle en bâton de hockey 119
6.6.2 Estimation de la concentration sans effet par régression 119

Section 7
Tests d'hypothèse(s) pour déterminer la concentration sans effet observé (CSEO) et la concentration avec effet minimal observé (CEMO) 122
7.1 Pertinence générale pour les essais d'écotoxicité 122
7.1.1 Essais à concentration unique 122
7.1.2 Essais à plusieurs concentrations 122
7.1.3 Expression des résultats sous forme de seuil 124
7.2 Particularités du plan d'expérience dans le test d'hypothèse(s) 124
7.2.1 Répétitions et unités expérimentales 125
7.2.2 Erreurs α et β 125
7.2.3 Puissance d'un essai de toxicité 126
7.2.4 Différence significative minimale 127
7.2.5 Bioéquivalence 128
7.2.6 Emploi des techniques sur les données quantiques 129
7.3 Préparatifs du test par analyse de variance 129
7.3.1 Tests de normalité et de comparaison de variances 130
7.3.2 Décisions après le test de distribution données 132
7.4 Analyse de variance 133
7.5 Tests de comparaisons multiples 134
7.5.1 Tests paramétriques 134
7.5.2 Tests non paramétriques 135

Section 8
Essais de mesure d'un double effet 138
8.1 L'effet quantique 138
8.2 La « croissance » en tant qu'effet sublétal 139
8.2.1 Options de mesure 139
8.2.2 Aspects conceptuels des options 141
8.2.3 Aspects statistiques des options 142
8.3 Le nombre de descendants en tant qu'effet sublétal 142
8.3.1 Interrelation entre la mortalité et la reproduction 143
8.3.2 Analyse séparée de la reproduction 144
8.4 Résumé et recommandations 145

Section 9
Quelques concepts et outils de statistique 146
9.1 Distributions normales et binomiales 146
9.1.1 Courbes normales 146
9.1.2 Distributions binomiales 146
9.2 Échantillons et populations 148
9.3 Signification statistique par opposition à signification biologique 149
9.4 Régression inverse 151
9.5 Différences significatives entre les CE 50 152
9.5.1 Paires de CE 50 153
9.5.2 Comparaison de CE 50 multiples 155
9.6 Différences significatives entre les CI p 155
9.6.1 Paires de concentrations inhibitrices (CI p) 156
9.6.2 Comparaison de CI p multiples 157

Section 10
Quand les résultats sont « difficiles » 158
10.1 Variabilité 158
10.2 Observations aberrantes 158
10.2.1 Vérification des erreurs et des modes opératoires 159
10.2.2 Modèles de rechange 159
10.2.3 Critères applicables aux observations aberrantes 160
10.2.4 Interventions à signaler 162
10.3 L'hormèse — stimulation à faibles concentrations 163
10.3.1 Les difficultés 164
10.3.2 Prise en considération des effets hormétiques dans la régression 166
10.3.3 Options face à l'hormèse 166
10.4 Relations concentration-effet déviantes 168
10.5 Interactions du mode opératoire sur les résultats 174

Références 177

Annexe A
Méthodes d'essai biologique et guides à l'appui, publiés par la Section de l'élaboration et de l'application des méthodes d'Environnement Canada A-188

Annexe B
Composition du Groupe intergouvernemental sur l'écotoxicité (en janvier 2004) . . B-190

Annexe C
Administration centrale et bureaux régionaux d'Environnement Canada C-192

Annexe D
Calculs employant des concentrations arithmétiques et logarithmiques D-193

Annexe E
La randomisation E-195

Annexe F
Calcul de la moyenne et des limites sur une carte de contrôle F-199

Annexe G
Tests s'appliquant aux résultats d'essai à concentration unique, sans répétition . . . G-203

Annexe H
Explication de la notion de *probit* et de la transformation log-probit H-207

Annexe I
Papier log-probabilité (ou log-probit) vierge I-211

Annexe J
Avantages et explication des logits J-213

Annexe K
La méthode de Spearman-Kärber K-216

Annexe L
Renseignements de base sur d'autres méthodes applicables aux données quantiques L-221

Annexe M
Méthodes non linéaires et méthodes du noyau applicables aux données quantiques M-223

Annexe N
Estimations ponctuelles applicables aux données quantitatives par lissage et interpolation N-225

Annexe O
Estimation des CI p par régression linéaire et non linéaire O-229

Annexe P
Test d'hypothèse(s) P-245

Annexe Q
Différences statistiques entre les CE 50 . . . . . Q-260

Annexe R
Médiane et quartiles R-262

Liste des tableaux

1. — Exemples de corrections apportées par la formule d'Abbott à divers effets observés chez les témoins dans un essai de toxicité quantique. 52
2. — Quatre exemples d'ensembles de données quantiques pour des essais de toxicité aiguë. 63
3. — Quatre exemples d'ensembles de données quantiques avec quelques effets partiels. . 67
4. — Types d'erreur dans les tests d'hypothèses et probabilités associées 125
5. — Différences significatives minimales (DSM) recommandées par l'USEPA pour des effets sublétaux manifestés dans certains essais de toxicité 129

Liste des figures

1. — Organigramme des principales catégories d'essais d'écotoxicité traitées dans le présent document. 5
2. — Carte de contrôle pour les essais avec un toxique de référence. 28
3. — Organigramme des méthodes statistiques applicables aux résultats de diverses catégories d'essais à concentration unique. 35
4. — Organigramme des méthodes d'analyse s'appliquant aux résultats des essais quantiques. 42
5. — Ajustement des droites des probits à vue d'œil à des ensembles représentatifs de données. 48
6. — Résultats des corrections apportées au moyen de la formule d'Abbott aux résultats d'un essai quantique, pour tenir compte de l'effet observé chez les témoins. 53
7. — Élargissement de l'intervalle de confiance des concentrations efficaces autres que la CE 50. 56
8. — Aspect graphique des régressions probit correspondant aux exemples A à D du tableau 2. 64
9. — Graphiques de données quantiques comportant quelques effets partiels (tableau 3). 68
10. — Démonstration graphique des transformations en probits et en logits 71
11. — Mortalité, en fonction du temps, de l'omble de fontaine exposée à de faibles concentrations d'oxygène dissous 81
12. — Temps d'effet médian chez le saumon de l'Atlantique exposé au cuivre et au zinc . 82
13. — Courbes de toxicité de deux toxiques hypothétiques. 85
14. — Inadaptation de la courbe de toxicité sur un graphique employant des échelles arithmétiques. 86
15. — Organigramme de l'analyse des résultats des essais de toxicité quantitatifs à plusieurs concentrations. 94
16. — Organigramme général de la sélection du modèle le plus approprié et de l'analyse statistique des données sur la toxicité quantitative 112
17. — Effet du cadmium sur l'inhibition de la croissance des frondes chez *Lemna minor* 115
18. — Exemples de régression en bâton de hockey 120
19. — Organigramme des analyses statistiques pour les tests d'hypothèses dans les essais de toxicité. 131

20. — Distributions normales. 147
21. — Distributions binomiales. 148
22. — Exemples d'observations peut-être aberrantes dans des essais de mesure de la croissance, au 7e jour, de larves de têtes-de-boule. 162
23. — Exemple de stimulation à faible concentration. 165
24. — Exemple de bonne relation linéaire entre la concentration et l'effet. 168
25. — Autre exemple d'une bonne relation entre la concentration et l'effet. 169
26. — Relation à pente raide entre le poids des larves de têtes-de-boule et les concentrations d'un effluent auxquelles elles sont exposées. 170
27. — Absence d'effet aux fortes concentrations avec anomalie à une concentration intermédiaire. 171
28. — Absence d'effet, apparemment anormale, à une concentration intermédiaire. 172
29. — Effet apparemment petit, mais variant à peine en fonction de la concentration. 173
30. — Exemple de performances améliorées en fonction de la concentration. 174
31. — Résultats d'un essai ne montrant que de grands effets. 175

Abréviations et symboles

α	alpha (lettre grecque)
β	bêta (lettre grecque)
σ	écart type de la population
Σ	sigma majuscule (lettre grecque)
χ, χ^2	khi (lettre grecque), khi-deux
ACLAE	Association canadienne des laboratoires d'analyse environnementale
APHA	American Public Health Association
ASTM	American Society for Testing and Materials
BPL	bonnes pratiques de laboratoire
C. V.	coefficient de variation
CCMRE	Conseil canadien des ministres des ressources et de l'environnement
CE 50	concentration efficace pour 50 % des individus, concentration efficace 50
CEMO	concentration avec effet minimal observé
CE p	concentration efficace pour *p* % d'individus
CESO	concentration avec effet de seuil observé
CI	concentration inhibitrice
CI p	concentration inhibitrice provoquant une diminution de *p* % des performances des sujets par rapport à celles du témoin
CL 50	concentration létale médiane, concentration létale 50
Comp.	comparer
CSEO	concentration sans effet observé
DSM	différence significative minimale
EC	Environnement Canada
EMV	estimation du maximum de vraisemblance
g	gramme
GLIM	modèle linéaire généralisé
GLM	modèle linéaire général
h	heure
ISO	Organisation internationale de normalisation
j	jour
kg	kilogramme
L	litre
mg	milligramme
OCDE	Organisation de coopération et de développement économiques
par ex.	par exemple
ppds	test de la plus petite différence significative de R.A. Fisher
R^2	coefficient de détermination
s	seconde
s	écart type (de l'échantillon)
S.-K.	méthode de Spearman-Kärber d'analyse applicable aux essais de toxicité quantique
s^2	variance
SPE	Service de la protection de l'environnement
Syn.	synonyme
$s_{\bar{x}}$	erreur type
USEPA	Agence de protection de l'environnement des États-Unis
V.	voir
‰	pour mille, millième
§	paragraphe

Glossaire

Toutes les définitions ci-après s'inscrivent dans le contexte du présent rapport. Elles pourraient ne pas être adaptées à d'autres contextes. Dans une définition, les mots ou expressions en italique sont l'objet d'une définition séparée. Dans certains cas, l'italique sert à attirer l'attention du lecteur.

Verbes auxiliaires

L'auxiliaire *doit* (*doivent*) exprime l'obligation absolue.

L'auxiliaire *devrait* (*devraient*) et le conditionnel d'obligation (*il faudrait*) expriment une recommandation ou la nécessité de respecter, dans la mesure du possible, la condition ou la marche à suivre.

L'auxiliaire *peut* (*peuvent*) exprime l'autorisation ou la capacité d'accomplir une action.

L'auxiliaire *pourrait* (*pourraient*) indique la possibilité ou l'éventualité.

Termes techniques

à renouvellement continu, Se dit d'un essai dans lequel on renouvelle continuellement la solution du milieu expérimental par l'apport constant ou intermittent, mais fréquent, de solutions fraîches.

à renouvellement intermittent, Se dit d'un essai de toxicité en milieu aquatique pendant lequel on renouvelle périodiquement la solution, habituellement au début de chaque période de 24 h. (Syn. *à renouvellement périodique.*)

aberrant, se dit d'une observation extrême, d'une mesure qui ne semble pas cadrer avec les autres résultats d'un essai.

aigu, Qui se manifeste dans une courte période (en secondes, en minutes, en heures ou en quelques jours), relativement à la longévité de l'organisme exposé.

ajout dosé (V. *enrichissement.*)

algorithme, Suite ordonnée de procédures permettant la résolution d'un problème. Ensemble de règles pour résoudre un problème. De façon générale, le mot a désigné, par le passé, des systèmes arithmétiques. Aujourd'hui, on l'emploie surtout dans le contexte de la résolution de problèmes mathématiques avec un ordinateur.

alpha (α), Seuil de signification fixé par l'expérimentateur, d'habitude à 0,05, soit la probabilité de 1/20, ce qui signifie qu'une différence de l'amplitude observée d'un phénomène pourrait, par l'effet du hasard, survenir dans de tels ensembles de données une fois sur 20. α est utilisé en analyse statistique, par ex. en *régression linéaire,* où il symbolise l'ordonnée à l'origine, et dans d'autres domaines. Le contexte aidant, ces autres utilisations se passent d'explications. (Voir *seuil de signification* et *erreur α.*)

ambiant, Qui entoure, qui environne, comme dans l'exemple suivant : « Les concentrations ambiantes en milieu de travail étaient de *x*... », ce qui signifie les concentrations dans l'air. Récemment, on a observé un usage souvent abusif du mot (comme dans « l'environnement ambiant »), et le meilleur remède consiste à ne pas l'employer.

analyse de covariance (ANCOVA), Technique d'évaluation des données obtenues par un plan d'expérience, dont les variables indépendantes sont tant continues que discrètes. On estime les différences significatives de la variable à laquelle on s'intéresse le plus en maintenant statistiquement l'autre variable constante. Un exemple pourrait être une régression simultanée du taux de survie (l'effet) sur la concentration de toxique (la variable continue), chez deux espèces de daphnies (la variable discrète). Si on s'intéresse avant tout à la relation entre l'effet et la concentration, on pourrait se servir de l'analyse de covariance pour l'estimer, en maintenant constant l'effet de l'espèce.

analyse de variance (ANOVA), Technique mathématique visant à déterminer méthodiquement s'il existe une différence significative entre les moyennes ou les variances d'échantillons découlant de différents *traitements*. Les traitements communs seraient l'exposition à différentes concentrations de toxique, y compris un témoin, ou l'emplacement dans différentes parties d'un panache d'effluent, par ex. la partie témoin, la partie proche de la zone de rejet et la partie éloignée. En analyse de variance, les variations d'arrière-plan entre les échantillons servent à affirmer s'il existe ou non des différences globales entre les traitements, mais elles ne permettent pas de dire lequel ou lesquels diffèrent des autres. En conséquence, on utilise souvent l'analyse de variance avant le test de comparaisons multiples. (V. le § 7.4.)

analyse non paramétrique, technique statistique ne présupposant pas une distribution sous-jacente des données. Elle n'emploie pas les *paramètres* (tels que la moyenne et la variance) de la *population* d'où les échantillons sont tirés. Le test non paramétrique tire des conclusions de la population, mais non des paramètres de la population. (V. *analyse paramétrique.*)

analyse paramétrique, Méthode de biostatistique tenant compte des paramètres de la population d'où les échantillons ont été tirés. D'habitude cela signifie que si on compare deux collections d'échantillons, les deux populations dont elles proviennent doivent suivre une loi normale et avoir des variances égales. Les échantillons analysés doivent posséder les mêmes caractéristiques que celles que l'on attribue par hypothèse à la population. (V. *analyse non paramétrique.*)

ANCOVA, (V. *analyse de covariance.*)

ANOVA (V. *analyse de variance.*)

assurance qualité (*AQ, assurance de la qualité*), Programme appliqué à l'intérieur d'un laboratoire pour que les travaux scientifiques et techniques arrivent à des résultats précis et exacts. Comprend la sélection des bonnes marches à suivre, la collecte des échantillons, le choix des limites, l'évaluation des données, la *maîtrise de la qualité,* ainsi que les compétences et la formation du personnel.

asymétrie, Défaut de symétrie de la courbe de fréquence d'une distribution de données. La courbe normale classique est symétrique, c'est-à-dire que ses branches de gauche et de droite sont les images inversées l'une de l'autre par rapport à la moyenne et que la médiane se confond avec la moyenne. Dans une courbe asymétrique à droite, la branche de droite est étirée, et la moyenne est supérieure à la médiane. Si on trace la courbe cumulative, on remarque que sa partie supérieure s'étire vers la droite en formant une courbe large. (V. le § 9.1 et l'annexe H.1.)

asymptotique (V. *seuil.*)

bêta (β), Probabilité de commettre une erreur bêta (conclure à un « faux négatif », c'est-à-dire à l'absence de différence significative quand il en existe effectivement une). Il existe une relation entre β et la puissance d'un test. β symbolise également un *paramètre* de population dans la formule de régression, dans laquelle il représente la pente. (V. *erreur bêta* et *régression linéaire.*)

biais, Erreur systématique entraînant une différence prévisible entre les estimations et leur valeur vraie (mais inconnue).

Par exemple, une piètre qualité de l'eau pourrait influer sur (biaiser) les résultats d'essais de toxicité en faisant croire à une toxicité apparente plus grande. (V. *exactitude* et *précision*.)

binaire, Équivalent à quantique. Une information binaire est en tout ou rien ; une observation faite sur une unité expérimentale individuelle doit prendre l'une de deux valeurs possibles. Une semence germe ou ne germe pas, etc.

bloc, Sous-ensemble (ou totalité) des traitements auxquels sont soumis les sujets de l'expérience. Chaque bloc est soumis aux mêmes traitements. Par exemple, un ensemble d'essais effectués dans un phytotron pourrait représenter un bloc, dans le dessein d'éliminer une cause de variabilité, à savoir la possibilité d'existence de différentes conditions accessoires dans l'ensemble d'essais, par suite des conditions existant dans différentes enceintes. Dans les essais de toxicité d'Environnement Canada, on insiste peu sur la constitution de blocs, parce que les modes opératoires des essais sont rigoureusement décrits, c'est-à-dire que l'on insiste sur la réduction des variations provenant de l'extérieur grâce à un plan d'expérience et à la maîtrise de l'appareillage et de la préparation des essais. (V. *répétition.*)

bonnes pratiques de laboratoire (BPL), Ensemble de normes régissant le plan d'expérience, la collecte des données et la conduite des études scientifiques et techniques dans le laboratoire. Le Conseil canadien des normes et Environnement Canada (EC) se sont dotés de programmes de BPL. Des normes sont également publiées par l'OCDE et l'USEPA.

carte de contrôle, Graphique de l'évolution des paramètres de toxicité d'un *toxique de référence.* La date de l'essai se trouve sur l'axe horizontal, tandis que sur l'axe logarithmique vertical on porte la concentration à laquelle l'effet est observé.

CE 50 (*concentration efficace médiane, concentration efficace à* 50 %, *concentration efficace* 50), Concentration de matière dans l'eau (par ex. en mg/L), un sol ou un sédiment (par ex. en mg/kg) que l'on estime causer un effet toxique spécifié chez 50 % des organismes en expérience. Dans la plupart des cas, la CE 50 et ses limites de confiance à 95 % résultent de l'analyse statistique des pourcentages d'organismes présentant l'effet spécifié à diverses concentrations expérimentales, après une période fixe d'exposition. La durée d'exposition doit être précisée (par ex. CE 50 après 72 h). La CE 50 décrit des effets quantiques, létaux ou sublétaux et elle ne s'applique pas aux effets quantitatifs (V. *CI p*). On pourrait utiliser d'autres pourcentages que 50 % (V. *CE p*).

CE p, Cette notion ne diffère de celle de *CE 50* que par la valeur de *p,* qui peut représenter tout pourcentage et qu'il faut préciser pour tout essai ou toute circonstance particulière. Des chercheurs et des organismes, particulièrement européens et internationaux, ont confondu *CE p* et *CI p,* mais il importe de continuer à distinguer ces notions.

CEMO (concentration avec effet minimal observé), La plus faible des concentrations de matière ayant un effet différent de l'effet observé chez les témoins, d'après les tests d'analyse statistique. (V. *CSEO,* la *concentration sans effet observé*.) [*O* ne signifie pas « observable », comme on l'écrit souvent à tort. La CEMO correspond à un effet que l'expérimentateur a effectivement observé. Un effet à une concentration inférieure pourrait avoir été *observable,* si on avait pu disposer d'une expérience plus puissante, si on avait consacré plus de temps à l'examen des organismes, si le microscope avait été plus puissant, etc. On ne devrait pas non plus intégrer le qualificatif « nocif » à l'expression qualifiant la concentration (concentration sans effet nocif observé ou CSENO). On devrait laisser à l'expérimentateur la possibilité de qualifier l'effet, sans imposer une définition extérieure de « nocif ».]

CESO (concentration avec effet de seuil observé), Valeur située quelque part entre la concentration sans effet observé (CSEO) et la concentration avec effet minimal observé (CEMO), étant la moyenne géométrique de ces deux concentrations. Elle présente l'avantage de remplacer ces deux estimations par une seule.

chronique, Qui survient pendant une période relativement longue d'exposition, qui représente habituellement une

proportion importante de la longévité de l'organisme, par ex. 10 % ou plus. En écotoxicologie, le mot s'est galvaudé, pour signifier *sublétal* ou, parfois, *couvrant le cycle vital,* mais cela ne devrait pas être. On devrait conserver à cet adjectif le sens qu'on lui attribue dans les autres domaines de la toxicologie, et, dans les autres situations, on devrait employer la terminologie convenable (*sublétal,* etc.).

CI p (*concentration inhibitrice p, concentration inhibitrice à tant pour cent*), Concentration correspondant à un pourcentage (désigné par *p*) d'effet. C'est une estimation ponctuelle de la concentration de la matière à l'étude, que l'on estime causer un pourcentage désigné d'inhibition d'une fonction biologique quantitative telle que la taille atteinte par les organismes au bout d'une période de croissance. Par exemple, la CI 25 du poids des organismes serait la concentration que l'on estime réduire le poids sec des organismes de 25 % par rapport au poids atteint par les organismes témoins. On devrait utiliser l'expression pour tout essai toxicologique permettant de mesurer un effet quantitatif ou une modification quantitative telle que la taille, le rendement de la reproduction ou la respiration. Pour ces essais quantitatifs, l'expression « CE 50 » (V. cette expression) ne convient pas. On peut estimer la CI p par régression ou, si nécessaire, par la méthode de *lissage* et d'*interpolation* à l'aide du programme informatique ICPIN.

CL 50 (*concentration létale médiane, concentration létale* 50 *%, concentration létale, concentration létale* 50), Concentration de matière dans l'eau, le sol ou un sédiment, que l'on estime mortelle pour la moitié des organismes en expérience. La CL 50 et ses limites de confiance à 95 % se calculent habituellement par l'analyse statistique des pourcentages de mortalités observés à plusieurs concentrations, après une période fixe d'exposition, qu'il faut préciser (par ex. CL 50 après 48 h). On pourrait spécifier d'autres taux de mortalité, par ex. la CL 20.

CMAT (*concentration maximale acceptable de toxique*), Notion définie, de manière diverse et discordante, mais actuellement considérée, de façon générale, comme synonyme de *concentration avec effet de seuil observé* (CESO), cette dernière expression étant celle que nous recommandons ici.

codage, Transformation des mesures originelles en nombres ou en symboles favorisant l'analyse ultérieure. On peut, pour cela, utiliser une simple opération arithmétique pour obtenir des valeurs plus maniables. Par exemple, on pourrait soustraire 840 de chaque élément de la série 842, 846, 849, 845..., ce qui donnerait, respectivement, 2, 6, 9, 5... Dans cet exemple, la moyenne calculée serait également inférieure de 840 à celle des données originelles. Par codage, on pourrait aussi représenter des catégories, par ex. en attribuant le code 1 aux femelles et le code 2 aux mâles.

coefficient de corrélation, À proprement parler, *coefficient de corrélation multiple.* (V. *R.*)

coefficient de corrélation multiple (V. *R.*)

coefficient de détermination (V. R^2.)

coefficient de détermination multiple (V. R^2.)

coefficient de variation (C.V.), Quotient de l'écart type divisé par la moyenne, d'habitude exprimé en pourcentage.

colinéarité, Corrélation entre des variables indépendantes. *multicolinéarité* possède la même signification. Si deux variables indépendantes ou explicatives sont fortement corrélées, la deuxième ajoute peu à l'explication de l'effet. Une colinéarité forte peut gonfler la variance des coefficients de régression partielle. Une colinéarité très forte peut empêcher l'inversion de la matrice dont on a besoin pour l'estimation de paramètres. On pourrait déceler la colinéarité comme suit, selon le cas : (1) création d'une matrice de corrélation des variables indépendantes et examen de cette matrice pour y déceler les corrélations fortes ; (2) examen des signes et de la grandeur des coefficients de régression, pour s'assurer qu'ils ont du sens. (V. *régression linéaire.*)

concentration avec effet de seuil observé (V. *CESO.*)

concentration avec effet minimal observé (V. *CEMO.*)

concentration d'essai (Syn. *concentration expérimentale*)

concentration-effet (V. *dose-réponse.*)

concentration efficace (V. *CE* 50).

concentration efficace à p % (V. *CE p*)

concentration efficace 50 (V. *CE* 50.)

concentration expérimentale (Syn. *concentration d'essai*)

concentration inhibitrice p (V. *CI p.*)

concentration « *inoffensive* », Concentration de la substance d'essai qui, estime-t-on, permet aux organismes de vivre et de se reproduire normalement dans leur habitat naturel. Il s'agit d'une notion de biologique et non la concentration estimée de façon statistique à la faveur d'une expérience. D'habitude, on guillemette l'adjectif, pour marquer qu'il n'est pas sûr que la concentration soit complètement inoffensive. (V. *concentration sans effet*)

concentration létale (V. *CL* 50.)

concentration létale 50 (V. *CL* 50)

concentration maximale acceptable de toxique (V. *CMAT*)

concentration sans effet (V. *CSE.*)

concentration sans effet observé (V. *CSEO.*)

confusion, Manifestation de l'influence d'une variable indésirable sur les résultats de l'expérience, d'une façon non aléatoire. Par exemple, si toutes les répétitions d'une concentration donnée étaient placées ensemble de façon à former un groupe régulier et séquentiel, dans le tableau des enceintes expérimentales, l'emplacement de ces répétitions dans le laboratoire se confondrait avec la concentration expérimentale.

contaminant, Matière biologique, chimique ou autre, ajoutée à un milieu naturel tel que l'air, l'eau, le sol ou un sédiment, directement ou non, du fait de l'activité humaine. Décelable expérimentalement, il pourrait entraîner des modifications chimiques ou physiques dans le milieu, mais il pourrait ne pas causer d'effet biologique néfaste. Habituellement, le terme s'applique aux matières en faible concentration, sans que des effets biologiques néfastes aient été prouvés. Divers organismes attribuent au mot *contaminant* des significations particulières, auxquelles s'ajoutent les significations découlant de certaines définitions ou de certains règlements nationaux et internationaux.

contamination, Processus par lequel un contaminant est introduit dans un milieu ou dans un être vivant ou résultat de cette introduction.

convergence, Propriété d'une série de nombres de tendre vers une limite définie ou un point commun.

corrélation, Rapport de proportionnalité entre deux variables, pas nécessairement en raison d'un lien de cause à effet. (V. *régression.*)

courbe de toxicité, Graphique des concentrations successives obtenues au cours d'un essai ou de plusieurs essais en fonction du temps, les deux sur des échelles logarithmiques (par ex. log de la CL 50 en fonction du log de la durée d'exposition). La courbe peut montrer si, au cours de l'essai, on a atteint un seuil de toxicité, c'est-à-dire une asymptote de la concentration indépendante du temps, ce qui est un élément important de connaissance de tout toxique (V. *CL* 50 *initiale*). La courbe de toxicité concerne habituellement les effets létaux, puisque, dans la plupart des essais de toxicité sublétale, les observations définitives des effets ne sont disponibles qu'à la fin de l'essai.

critère, Selon la définition du CCRME (1987), « donnée scientifique évaluée aux fins du calcul des seuils recommandés pour des utilisations particulières de l'eau. V. *ligne directrice relative à la qualité.* Un usage plus répandu, aux États-Unis et ailleurs, donne à *critère* la signification attribuée à *ligne directrice* dans le présent glossaire. Par exemple, Rand (1995) définit le critère de qualité de l'eau comme « une estimation, fondée sur des jugements scientifiques, de la concentration d'une substance ou d'un autre constituant dans l'eau qui, si elle n'est pas dépassée, protégera un organisme, une communauté d'organismes ou une utilisation ou une qualité prescrite de l'eau avec un degré convenable de sécurité ». Ces définitions qui concernent le domaine de l'eau concernent également d'autres milieux tels que le sol.

CSE (concentration sans effet), Concentration de toxique que l'on pense n'avoir aucun effet sur un organisme donné. La CSE est un peu une notion idéalisée. Il faut l'estimer ou en prévoir la valeur par modélisation ou extrapolation. Elle est analogue à un *paramètre* d'une *population*. Il faut la déduire des résultats d'un essai de toxicité plutôt que l'observer, parce que plus d'essais ou différentes sortes d'essais pourraient révéler des effets à des concentrations inférieures.

CSEO (concentration sans effet observé), Concentration immédiatement inférieure à la CEMO, parmi toutes les concentrations expérimentales. (C'est presque toujours, également, la concentration expérimentale maximale dont l'effet sur les organismes n'est pas différent de l'effet sur les organismes témoins, d'après le test d'analyse statistique. Il est possible, cependant, qu'une réaction irrégulière n'aboutisse à aucun effet significatif à une concentration supérieure à la CEMO. La définition donnée à la CSEO permet d'éviter cela.)

DE 50 (*dose efficace médiane, dose efficace à* 50 %, *dose efficace* 50), C'est *mutatis mutandis* la CE 50, sauf qu'il s'agit d'une dose toxique.

degré de liberté, Caractéristique d'un ensemble de données soumises à l'analyse statistique. C'est un concept de statistique énonçant le nombre de degrés de liberté avec lesquels on peut spécifier une valeur. Par exemple avec n observations et une moyenne fixée, toute valeur peut être choisie pour $n - 1$ observations. Cependant, la dernière observation est fixée par la moyenne et les valeurs des $n - 1$ premières observations. Le nombre de degrés de liberté est de $n - 1$. On se sert souvent des degrés de liberté pour estimer une variance moyenne ou la moyenne des carrés des erreurs.

dérivée (V. *dérivée partielle.*)

dérivée partielle, Notion ayant rapport aux variables indépendantes d'une fonction. On peut l'expliquer à partir d'une fonction très simple telle que $Y = aX$. Y est la *variable* dépendante, X la *variable* indépendante et a un *paramètre.* La dérivée est la variation de Y par rapport à la variation de X (c'est-à-dire la pente). Il s'ensuit que la dérivée est $\delta Y/\delta X = a$. Si, cependant, la fonction possède au moins deux variables indépendantes, il faut la dériver pour chacune de ces dernières afin de décrire la pente. Par exemple, si la fonction est $Y = aX_1 + bX_2$, elle possède deux *dérivées partielles,* à savoir $\delta Y/\delta X_1 = a$ et $\delta Y/\delta X_2 = b$.

différence significative minimale (DSM), Différence dans les mesures, qui devrait exister entre la concentration témoin et une concentration d'essai, pour conclure qu'il existe un effet significatif à cette concentration, d'après le test statistique utilisé.

distribution, Répartition d'une caractéristique parmi les membres d'une classe, souvent représentée graphiquement par une courbe. Dans l'usage courant, *distribution* est synonyme de *distribution de probabilité,* c'est-à-dire la fréquence relative des valeurs que peut prendre une variable. Par exemple, dans l'essai de reproduction de daphnies, le nombre moyen de nouveau-nés par adulte est d'habitude dans la fourchette de 18 à 22. La fréquence relative des valeurs dans cette fourchette est beaucoup plus grande que celle d'une valeur comme, disons, 35. La distribution de probabilité décrit ces fréquences relatives. On peut s'en servir pour déterminer la probabilité de survenue d'une observation ou d'un ensemble d'observations pour une distribution donnée.

distribution binomiale ou *distribution de probabilité binaire*, Probabilité qu'une variable aléatoire binomiale soit représentée par une valeur spécifiée. On peut se la représenter comme une courbe montrant la répartition des fréquences associées aux proportions d'un phénomène quantique positif (par ex. la mortalité, dans un essai de toxicité). Les fréquences dépendent du nombre d'observations et de la probabilité (p) de survenue du phénomène. Pour les tailles d'échantillon moyennes (disons de 25 unités) ou plus grandes, associées à $p \approx 0{,}5$, la distribution binomiale ressemble à la *distribution normale* bien connue, en forme de cloche. Dans une telle distribution, beaucoup d'observations se regrouperaient près de la proportion de 0,5, en étant de moins en moins nombreuses à mesure que les proportions s'écarteraient de cette valeur pour tendre vers 0 ou 1,0. (V. *distribution de probabilité*.)

distribution de Gompertz (V. *distribution de Weibull.*)

distribution de Poisson, Distribution comportant des dénombrements d'un élément distribué au hasard, dans l'espace ou le temps. Un exemple serait le dénombrement des cellules algales sur un quadrillage. Si la probabilité était faible (mais constante) et que le nombre d'observations était grand, la distribution de Poisson tendrait vers la distribution *binomiale.*

distribution de probabilité, Fonction décrivant la probabilité qu'une variable aléatoire soit égale ou inférieure à une valeur non précisée. Un exemple bien connu est celui de la distribution normale en forme de cloche. Si la variable aléatoire est égale à 1,645, la probabilité qu'elle soit inférieure à cette valeur est de 95 %. (V. *distribution.*)

distribution de Weibull, Version généralisée d'un modèle exponentiel. On peut s'en servir pour des ajustements empiriques des données sur la relation entre la dose et l'effet. La distribution est sigmoïde, mais elle permet à la forme de la courbe de différer au-dessus et au-dessous du point d'inflexion, ce qui est un avantage par rapport aux distributions probit ou logit. Le modèle de Gompertz, qui équivaut essentiellement à celui de Weibull, est utile à la régression non linéaire (V. le § 6.5.8).

distribution des tolérances, Distribution, au sens statistique, des effets parmi les organismes exposés à une seule concentration d'agent toxique. On peut donner en exemple la croissance d'un groupe d'organismes exposés à une concentration donnée d'un toxique. Chaque organisme présentera une étendue d'effets. Il y aura un effet moyen sur la croissance, avec dispersion des individus par rapport à cette moyenne. Cette distribution par rapport à la moyenne est la *distribution des tolérances.* Si on expose un autre groupe à une concentration différente, l'effet moyen sera différent, mais on pose, par hypothèse, que la distribution des tolérances restera la même, c'est-à-dire qu'elle possédera la même variance. Cet usage statistique diffère de la définition établie de *tolérance* au sens biologique et toxicologique. Dans le présent document, nous avons évité d'employer l'expression « distribution des tolérances ».

distribution logistique, Fonction de distribution statistique qui s'est révélée utile dans les essais quantiques et les régressions de données quantitatives. (V. *logit.*)

distribution normale (*distribution gaussienne, distribution de probabilité normale, loi normale*), Série ordonnée et symétrique d'observations en forme de cloche. La série fait correspondre la fréquence à la valeur de la variable. Dans une distribution normale, la plupart des observations sont groupées autour de la valeur moyenne, avec de moins en moins d'observations vers les valeurs extrêmes. La forme de la courbe est déterminée par la moyenne et l'écart type, 68,3, 95,4 et 99,7 % des observations étant comprises dans l'intervalle de ± 1, 2 et 3 écarts types, respectivement, par rapport à la moyenne. Les courbes en forme de cloche ne sont pas toutes normales, et la normalité est définie par une équation particulière et complexe qui comprend la moyenne et l'écart type ainsi que les constantes π (3,14159) et *e* (la base des logarithmes népériens). La distribution normale joue un rôle primordial dans la théorie statistique, en raison de ses propriétés mathématiques. Il en est de même en biologie, parce que de nombreux phénomènes biologiques obéissent à la loi normale. Beaucoup de tests statistiques reposent sur l'hypothèse de la distribution normale des données et, en conséquence, il peut être nécessaire de vérifier si cela est vrai pour un ensemble particulier de données.

DL 50 (*dose létale médiane, dose létale* 50*, dose létale* 50 %), Notion possédant la même définition que celle de la CL 50, sauf que, au lieu de concentration(s), on parle de *dose*(s).

donnée, Fait représenté sous une forme numérique, observation ou élément d'information numérique.

dose, Quantité d'une substance ou d'un toxique ayant pénétré dans l'organisme en expérience. La dose est inconnue dans la plupart des essais d'écotoxicité, par lesquels on estime l'effet de concentrations dans le milieu. (V. *dose-réponse.*)

dose efficace 50 (V. *DE* 50)

dose létale 50 (V. *DL* 50)

dose létale médiane (V. *DL* 50.)

dose-réponse, Locution adjective invariable s'appliquant aux notions classiques de la pharmacologie ou de la toxicologie telles que la « relation dose-réponse », la gamme des modifications observées chez les organismes en relation avec la quantité de médicament ou de toxique. L'expression sert de façon très générale en écotoxicologie, bien que la locution *concentration-effet* soit habituellement plus appropriée. Comme nous l'avons mentionné déjà, la plupart des essais d'écotoxicité concernent des concentrations ambiantes, plutôt que des doses dans les organismes. De même, le mot *réponse* convient en médecine ou en pharmacologie, où l'être humain ou tout autre organisme peut présenter une amélioration apportée par une dose de médicament, tandis qu'en toxicologie l'organisme souffre davantage de l'effet du toxique qu'il ne répond à ce dernier.

DSM (V. *différence significative minimale.*)

eau de porosité, Eau occupant les interstices entre les particules d'un sédiment. La quantité d'eau de porosité s'exprime en pourcentage en poids du sédiment humide.

écart type, Mesure de la dispersion des observations faites sur un échantillon, par rapport à la valeur moyenne de cet échantillon. Égale la racine carrée de la *variance* et, par définition, ce ne peut être qu'un nombre positif. On le symbolise par *s*.

échantillon, Sous-ensemble d'une population, choisi pour obtenir des renseignements sur cette *population* et effectuer des inférences sur cette dernière. En conséquence, il importe de définir clairement la population à laquelle on s'intéresse et d'en prélever un échantillon représentatif ; cela se fait souvent par *échantillonnage au hasard.*

échantillon aléatoire, Échantillon dans lequel les individus (ou éléments) d'une *population* ont une probabilité égale d'être tirés pour en faire partie. La plupart des techniques statistiques exigent un échantillonnage aléatoire pour que les inférences soient valides.

échantillonnage au hasard (V. *échantillon aléatoire.*)

échantillons réitérés, Échantillons séparés de sol, de sédiment, etc., prélevés sur le terrain, au moyen de méthodes identiques et dans la même station. Par définition, chaque échantillon réitéré est assujetti au même traitement. Il s'agit de parvenir à une évaluation plus représentative de la qualité du substrat échantillonné et d'estimer la variation de la qualité et/ou la variation de l'échantillonnage du substrat. Les échantillons réitérés doivent êtres gardés dans des récipients séparés. À partir de ces échantillons, on pourrait obtenir des répétitions de chaque concentration utilisée dans un essai de toxicité ; cela est souvent préconisé dans les essais sur les sols ou les sédiments. Ces répétitions de l'essai constitueraient les véritables *échantillons réitérés de terrain*, de sorte que l'essai évaluerait la variation de la matière à l'étude et la variation de son échantillonnage, de même que toute variation entre les répétitions attribuable aux conditions existant dans le laboratoire. Dans un essai, les *répétitions de laboratoire* seraient deux répétitions ou plus de chaque traitement, obtenues par subdivision de l'échantillon de la matière ou par prélèvement d'un *sous-échantillon* de cet échantillon. Dans ce cas, l'essai ne donnerait qu'une indication de la variation due aux conditions existant dans le laboratoire : on ne doit absolument pas l'interpréter comme signe d'une variation dans la matière (disons, un sédiment lacustre) ou dans le prélèvement d'échantillons de cette matière. Dans les essais de toxicité, il est habituellement inutile de disposer de répétitions de laboratoire, et ces dernières ne sont pas recommandées, si ce n'est pour tenir compte de la taille des récipients ou pour quelque raison semblable. Ces répétitions pourraient cependant trouver quelque utilité pour la régression, afin de permettre de distinguer entre l'erreur de mesure d'un effet et l'écart réel de l'effet par rapport à la droite ajustée. Pour l'analyse chimique, on pourrait analyser des répétitions de laboratoire pour évaluer la *précision* des dosages.

échelle probit, Échelle dont la valeur centrale est de 5,0, qui représente l'effet médian prévu dans un essai de toxicité quantique (on s'attend à ce que 50 % des organismes subissent l'effet). La plupart du temps, une échelle de 3 à 7 probits conviendrait. Le 2 de cette échelle (probit 2) correspondrait à un effet prévu chez 0,1 % des organismes, le 3 chez 2,3 %, le 4 chez 16 %, le 6 chez 84 %, le 7 chez 97,7 % et le 8 chez 99,9 %. (V. *probit, variable normale équivalente* et le § 4.5.1.)

écotoxicologie, Comme l'écotoxicologie est une subdivision de la toxicologie, les deux possèdent la même définition générale. Cependant, l'écotoxicologie insiste sur les effets exercés sur les organismes sauvages et les communautés naturelles, sans exclure la sécurité de l'espèce humaine comme élément constitutif des écosystèmes.

effet, En toxicologie, modification biologique mesurable. Cette modification peut concerner une structure, la physiologie, le comportement, etc. Dans un essai de toxicité, on devrait l'estimer par comparaison à des mesures effectuées sur des organismes faisant partie du groupe témoin. L'analyse statistique considère généralement les degrés d'effet qui n'ont pas été mesurés chez les témoins et que l'on présume donc résulter de l'exposition aux constituants toxiques de la matière visée par l'essai.

effet gradué (V. *quantitatif.*)

effet partiel, Manifestation de l'effet chez une partie seulement des organismes dans une enceinte expérimentale. La notion peut s'appliquer à un effet létal, comme la *mortalité partielle,* ce qui signifierait que certains organismes seulement sont morts.

effluent, Tout déchet liquide (par ex. industriel, urbain) rejeté dans l'environnement. Il n'est pas besoin d'utiliser l'expression *« effluent entier ».*

élutriat, Solution aqueuse obtenue après avoir ajouté de l'eau à une matière solide (par ex. sol, sédiment, stériles, boues de forage, déblais de dragage), avoir agité le mélange, puis après avoir récupéré le liquide par centrifugation, filtration ou décantation du surnageant.

élutriat de sol (V. *élutriat.*)

emboîté, Se dit d'un plan d'expérience dans lequel toutes les combinaisons possibles d'un facteur ne peuvent pas exister (comp. *croisé*). Si un essai tient compte du sexe de l'organisme et des concentrations de toxique comme facteurs, avec triple dosage des résidus du toxique dans les tissus, il est impossible de planifier une expérience dans laquelle chaque animal se trouve à chaque combinaison de facteurs. Les dosages en triple du résidu sont des sous-échantillons qui sont emboîtées dans le facteur « animal » (Syn. *niché.*)

en conditions statiques, Se dit d'un essai de toxicité en milieu aquatique pendant lequel on ne renouvelle pas les solutions.

enrichissement, Addition d'une quantité connue de substance ou de produit chimique à un sol ou à un sédiment. Habituellement, ce sol ou ce sédiment n'est pas contaminé ou c'est un sol ou sédiment témoin, mais, parfois, il peut être contaminé. La substance ajouté serait habituellement pure, mais ce pourrait être du sol ou du sédiment à l'étude. Après l'enrichissement, on homogénéise le mélange.

erreur, ‖ *Taux d'erreur se rapportant aux comparaisons deux à deux,* rapport du nombre d'inférences fausses au nombre total d'inférences effectuées. Le *taux d'erreur expérimentale* est la probabilité de faire au moins une erreur de première espèce (erreur α) dans toutes les comparaisons (relatives à un effet donné) au cours de l'expérience. Par exemple, dans le contexte d'une étude de la toxicité dans les sédiments, les comparaisons se feraient entre l'effet moyen de chaque emplacement avec la moyenne du témoin. « L'expérience » serait la totalité de l'étude. Le taux d'erreur n'engloberait pas les comparaisons relatives à tout autre effet biologique. (V. *erreur de première espèce* et *erreur de seconde espèce.*)

erreur alpha (*erreur α, erreur de première espèce*), Erreur consistant à rejeter une hypothèse nulle exacte. Autrement dit, on affirme qu'une différence est significative alors qu'elle ne l'est effectivement pas.

erreur bêta (*erreur β, erreur de seconde espèce*), Erreur consistant à ne pas rejeter une hypothèse nulle inexacte. Autrement dit, on affirme qu'une différence n'est pas significative alors qu'elle l'est effectivement.

erreur d'échantillonnage (V. *précision.*)

erreur de première espèce (V. *erreur alpha.*)

erreur de seconde espèce (V. *erreur bêta.*)

erreur expérimentale (V. *précision.*)

erreur systématique (V. *biais.*)

erreur type (*par rapport à la moyenne* [$s_{\bar{x}}$]), On peut calculer l'erreur type de toute statistique, par exemple l'estimation de la pente d'une régression linéaire. Cependant, l'utilisation la plus répandue de l'erreur type en toxicologie concerne la moyenne de l'échantillon. Cette erreur type se calcule comme l'*écart type* de l'échantillon divisée par la racine carrée du nombre d'observations dans l'échantillon. Cette erreur type calculée est une estimation de la dispersion que manifesteraient un certain nombre de valeurs moyennes, si ces moyennes représentaient un certain nombre

d'échantillons prélevés dans la même population. L'erreur type d'un groupe de moyennes est donc l'équivalent de l'écart type d'un groupe d'observations portant sur un seul échantillon. Dans la pratique, on estime l'erreur type à partir d'un seul échantillon, comme nous venons de l'affirmer.

essai biologique, Essai visant à estimer, à l'aide d'organismes vivants, la concentration ou la puissance d'une matière telle qu'un médicament. En pharmacologie, on estime habituellement la puissance en comparant les résultats à ceux d'une préparation étalon, éprouvée simultanément. On a aussi appliqué l'expression aux essais appliqués à l'environnement, mais *essai de toxicité* décrit plus spécifiquement ces essais et leurs objectifs et c'est l'expression préconisée.

essai couvrant le cycle vital, Essai au cours duquel on observe les organismes à partir d'un stade d'une génération jusqu'au même stade au moins dans la génération suivante.

essai de toxicité, Détermination de l'effet d'une matière sur un groupe d'organismes choisis, dans des conditions définies. L'essai d'écotoxicité permet habituellement de mesurer, selon le cas : *a*) la proportion d'organismes touchés (effet quantique) ; *b*) le degré d'effet manifesté (effet quantitatif) après exposition à des concentrations précises de substance chimique, d'effluent, d'élutriat, de percolat, de milieu récepteur, de sédiment ou de sol.

estimation de la toxicité (V. *paramètre de toxicité.*)

estimation ponctuelle, Nombre unique, calculé pour représenter un ensemble de données de données sur la toxicité ou censé représenter cet ensemble, par ex. la CE 50 ou la CI 25.

étendue, Différence entre la plus grande et la plus petite des valeurs d'un ensemble de données. On l'exprime habituellement en précisant ces valeurs.

exactitude, Qualité de l'accord entre la valeur mesurée (ou estimée) et la vraie valeur. Dans les essais de toxicité, on ne peut pas mesurer l'exactitude parce qu'il n'y aucun moyen de connaître la vraie valeur de la toxicité. Syn. *justesse.* (V. *précision.*)

exposant, Symbole ou chiffre placé à droite et un peu au-dessus d'une quantité pour désigner le nombre de fois que l'on doit multiplier cette quantité par elle-même. Par exemple $5^2 = 5 \times 5 = 25$. (V. *logarithme.*)

facteurs croisés (V. *factoriel.*)

factoriel, Se dit d'un plan d'expérience dans lequel existent toutes les combinaisons possibles de facteurs. Par exemple, dans le cas de deux facteurs (le sexe de l'organisme en expérience et la concentration de toxique) et un effet mesuré de résidu de toxique dans les tissus, on peut planifier une expérience dans laquelle chaque sexe est exposé à chaque concentration. Dans ce cas, les facteurs sont dits *croisés.* (V. *niché.*)

GLIM (V. *modèle linéaire généralisé.*)

GLM (V. *modèle linéaire général.*)

hétéroscédasticité, Hétérogénéité des résidus, que présentent les données dans un nuage de points (V. les fig. O.2B et O.2C de l'annexe O). Il y a hétéroscédasticité lorsque la variabilité des résidus diffère de façon significative de celle des variables indépendantes (c'est-à-dire les concentrations expérimentales ou les traitements). Dans l'analyse statistique et l'estimation des résidus (par le test de Levene, par exemple), si les données expérimentales présentent une hétéroscédasticité (c'est-à-dire que les résidus ne sont pas homogènes), c'est qu'il existe une différence

significative entre la variance des résidus aux différentes concentrations ou aux différents traitements. (V. *homoscédasticité* et *résidu.*)

homoscédasticité, Homogénéité des résidus, que présentent les données dans un nuage de points (V. la fig. O.2A de l'annexe O). Il y a homoscédasticité lorsque la variabilité des résidus ne diffère pas de façon significative de celle des variables indépendantes (c'est-à-dire les concentrations expérimentales ou les traitements). Dans l'analyse statistique et l'estimation des résidus (par le test de Levene, par exemple), si les données expérimentales présentent une homoscédasticité (c'est-à-dire que les résidus sont homogènes), c'est qu'il n'existe pas de différence significative entre la variance des résidus aux différentes concentrations ou aux différents traitements. (V. *hétéroscédasticité* et *résidu.*)

hormèse, Effet qui, aux faibles concentrations de la matière à l'étude, stimule les organismes en expérience, dont les performances sont meilleures que celles des organismes témoins. Aux concentrations supérieures, on constate des effets nuisibles. Une catégorie plus générale de stimulation à faible dose engloberait d'autres causes possibles de stimulation, par ex. l'effet des solvants, l'erreur expérimentale ou une « stimulation suffisante » parmi les organismes de laboratoire.

initial, En parlant de la CL 50 ou de la CE 50 pour des effets quantiques aigus, intensité du stimulus (c'est-à-dire la concentration) à laquelle on peut s'attendre à un effet chez (tout juste) 50 % des organismes exposés après une période indéfiniment longue. C'est la concentration qui serait tout juste suffisante pour agir sur l'organisme médian (l'organisme « typique » ou « moyen »). L'expression d'origine, plus générale et encore utile, est *concentration létale initiale* (Fry, 1947). Sont équivalentes les expressions « CE 50 seuil », « CE 50 indépendante du temps » et CE 50 asymptotique », qui font toutes allusion à la courbe de toxicité devenant parallèle à l'axe du temps. L'adjectif *initial* permet d'éviter les connotations contradictoires du mot « seuil ». La définition de *initial* devient plus arbitraire et plus difficile lorsque l'on parle d'effets quantitatifs sublétaux, auxquels manque le critère évident et usuel d'effet médian, utilisé pour les essais de toxicité quantique. Dans le cas des essais quantitatifs, on pourrait mieux définir *initial* comme la concentration la plus faible à laquelle s'est manifestée une modification nocive significative dans l'effet que l'on estimait (comme la croissance). Dans la pratique, cette estimation d'un effet quantitatif varierait selon le plan de l'expérience et la précision de l'essai.

intervalle interquartile (V. *quartile.*).

itération, Procédé mathématique servant à estimer les paramètres d'une régression (c'est-à-dire pour ajuster une courbe). Cela comporte des approximations successives des estimations à la faveur de cycles de calcul, chaque cycle se fondant sur l'approximation précédente et améliorant les estimations.

justesse, Syn. *exactitude.*

khi-deux (χ^2), Statistique d'un test servant parfois à estimer l'ajustement d'un modèle à un ensemble de données.

létal, Qui cause directement la mort. Celle-ci se définit habituellement par la cessation de tous les signes visibles de mouvement ou d'activité et par l'absence de ces signes en réaction à une stimulation externe légère.

ligne directrice (V. *ligne directrice relative à la qualité.*)

ligne directrice relative à la qualité, Limite numérique de concentration, reposant sur des bases scientifiques (on parle alors de *nombre-guide*) ou énoncé narratif, recommandé pour appuyer et préserver une utilisation désignée d'un milieu tels que le sol, l'air ou l'eau (« ligne directrice relative à la qualité des sols », etc.). L'*objectif de qualité* se définit de même, sauf qu'il s'applique à un emplacement précis. Des provinces ont établi des listes d'*objectifs relatifs à la qualité de l'eau*, qui reflètent des conditions officiellement visées. Une *norme de qualité* est un objectif reconnu

dans des lois et des règlements applicables à l'environnement par un pouvoir public.

limites de confiance, Leur valeur est si semblable à celle des *limites fiducielles* que l'on assimile l'une à l'autre les deux notions dans le présent document. Ces limites, en parlant d'une CE 50 ou d'une CI p, représentent les concentrations supérieure et inférieure entre lesquelles on pense que se trouve la valeur vraie, pour un niveau précisé de probabilité. Les limites de confiance au seuil de 95 % (limites de confiance à 95 %) signifient que 19 fois sur 20 la valeur vraie se situe à l'intérieur de l'intervalle ainsi spécifié.

limites de la zone de confiance, Limites calculées logarithmiquement, situées à ± 2 écarts types de part et d'autre de la moyenne géométrique « historique » des paramètres de toxicité d'un toxique de référence, sur la carte de contrôle, qui permet d'évaluer la variabilité des résultats des essais de toxicité effectués sur ce toxique.

limites fiducielles (V. *limites de confiance.*)

lixiviat, Eau, usée ou non, ayant traversé une épaisseur de sol ou de déchets solides.

log, Notation désignant le *logarithme* de base 10 ou logarithme décimal.

logarithme, Méthode de codage mathématique. Dans le cas du logarithme décimal, logarithme de base 10 ou log, c'est la puissance à laquelle il faut élever la base fixe de 10 pour obtenir le nombre représenté par le logarithme. Ainsi le logarithme 2 représenterait $10^2 = 100$, c'est-à-dire log de base 10 de 100 = 2 ou $\log_{10} 100 = 2$. D'autres exemples aideront à comprendre : $\log_{10} 700 = 2{,}84510$; $\log_{10} 70 = 1{,}84510$; $\log_{10} 7 = 0{,}84510$; $\log_{10} 0{,}7 = -0{,}15410$ (ou 9,84510 – 10). L'addition (ou la soustraction) de logarithmes équivaut à multiplier (ou à diviser) les nombres qu'ils représentent. (Voir *exposant.*) Dans l'expression dans $10^2 = 100$, l'exposant est 2, ce qui rattache la notion à celle des logarithmes. Le logarithme *népérien* ou *naturel* (ou *ln* comme dans ln 100 = 4,60517) est de base *e*, dont la valeur est de 2,71828... On peut utiliser l'un ou l'autre type de logarithme en toxicologie, pour autant que l'on soit cohérent du début à la fin d'un calcul. La base *e* est importante pour certains concepts mathématiques comme l'intérêt composé, la fonction exponentielle, la théorie des probabilités, les équations de croissance, etc.

logarithme naturel (V. *logarithme.*)

logarithme népérien (V. *logarithme.*)

logarithmique, Se dit d'une suite dont le logarithme de chaque nombre est supérieur d'une quantité constante au logarithme du nombre qui le précède. On pourrait aussi dire que ces nombres constituent une suite géométrique, puisque chacun serait supérieur à celui qui le précède par un multiplicateur constant.

logit, Déviation logistique équivalente. Il s'agit d'une transformation particulière des données, applicable à la proportion d'organismes subissant un effet dans un essai de toxicité quantique (binaire), qui aboutit d'habitude à un redressement de la courbe sigmoïde de l'effet. Pour obtenir le logit, on divise la proportion d'organismes touchés (p) à une concentration donnée par ($1 - p$). Le logarithme du résultat est le logit. V. le § 4.5.1 et *probit*, ainsi qu'une discussion approfondie de la notion dans l'annexe J. Les logits constituent également une façon utile d'ajuster une régression aux données quantitatives. Les résultats sont exprimés en proportions d'organismes ayant atteint les valeurs spécifiées de l'effet mesuré. On donne des exemples dans le § 6.5.8 avec de plus amples détails dans l'annexe O.

loi de Gompertz (V. *distribution de Weibull.*)

maîtrise de la qualité, Actions précises, englobées dans le programme d'assurance qualité : normalisation, étalonnage, répétitions, échantillons témoins et estimations statistiques des limites relatives aux données.

matière, Somme de toutes les *substances* qu'elle renferme. La matière possède des caractéristiques plus ou moins uniformes. Un sol, un sédiment ou une eau de surface sont des matières.

matrice, En pédologie et en sédimentologie, la substance fondamentale de l'échantillon, pour ce qui concerne ses effets physicochimiques, indépendamment des contaminants présents. Par ses caractéristiques, la matrice exerce sur les organismes en expérience un effet dit de matrice.

médiane, Valeur qui, dans un ensemble de données rangées dans l'ordre croissant (ou décroissant) de valeur, partage ces valeurs en deux effectifs égaux. On se trouve à diviser le nombre d'individus dans la série ainsi divisée et non les valeurs arithmétiques de ces individus. Si le nombre d'individus était impair, la médiane serait l'individu du milieu. Dans le cas d'un nombre pair d'individus, la médiane serait d'habitude la demi-somme des valeurs numériques des deux individus du milieu. Si, pour compliquer les choses, plus de deux individus du milieu ou près du milieu de la série avaient la même valeur, la médiane serait déterminée par interpolation, dans l'hypothèse selon laquelle les valeurs médianes seraient rangées symétriquement de part et d'autre de l'intervalle séparant individus en effectifs égaux. Les manuels de statistique donnent les formules appropriées de calcul de la médiane. Celle-ci renseigne moins que la moyenne, parce qu'elle ne tient pas compte de la valeur réelle de chaque mesure. Cependant, on peut être justifié de la choisir pour décrire la tendance centrale d'une population à distribution asymétrique, parce que les valeurs extrêmes n'influent pas autant sur la médiane que sur la moyenne (Zar, 1999). [V. *quartile.*]

mesure répétée, Répétition d'une observation numérique, à des dates différentes, sur la même unité expérimentale. Catégorie distincte des méthodes statistiques pour ce type d'observations, l'analyse des mesures répétées n'est pas traitée dans le présent document.

méthode de référence, Mode opératoire d'un essai de toxicité, constitué d'un ensemble explicite d'instructions et de conditions décrites avec précision dans un document écrit. Contrairement aux autres méthodes polyvalentes (générique) d'essai biologique publiées par Environnement Canada, l'emploi d'une méthode de référence est habituellement spécifiquement exigé par un règlement.

méthode des moindres carrés, Méthode d'ajustement d'une courbe à un ensemble de données. Elle permet de réduire au minimum la somme des carrés des écarts entre les valeurs observées et les valeurs prévues respectives.

méthode du maximum de vraisemblance, méthode mathématique d'estimation des paramètres d'une relation à laquelle on s'intéresse. Avec les estimations du maximum de vraisemblance (EMV), on tente d'estimer les paramètres qui feraient qu'il est le plus vraisemblable d'observer les données effectivement collectées (SPSS, 1996). Par exemple, les paramètres pourraient être la moyenne et la variance d'une distribution de données. « La *vraisemblance,* c'est-à-dire la probabilité d'observer des valeurs particulières pour un ensemble de paramètres, se définit comme une quantité proportionnelle à la probabilité que, si ces valeurs étaient les paramètres, la totalité des observations devrait constituer les données enregistrées » (traduction, d'après Finney, 1978, p. 58). Notion à ne pas confondre avec celle de la méthode des moindres carrés, ni celle du khi-deux minimal.

milieu, Dans les essais de toxicité, la matière entourant ou transportant les organismes, par ex., dans le cas des bactéries, le milieu de culture (bouillon nutritif ou substrat), l'eau dans laquelle nagent les poissons, le sol entourant les vers de terre.

milieu récepteur, Eau de surface (par ex. d'un cours d'eau, d'un lac ou d'une baie) dans laquelle on a rejeté un déchet ou qui est sur le point de recevoir un tel rejet (par ex. eau d'un cours d'eau juste en amont du point de rejet). Cette distinction doit être explicitée au moment opportun.

MMV (V. *méthode du maximum de vraisemblance.*)

modèle linéaire général (GLM), Catégorie de modèles (et non technique mathématique particulière) aux caractéristiques et à l'approche très semblables, dans lesquels une seule variable dépendante (pouvant faire l'objet de multiples mesures dans une unité expérimentale) est fonction d'une variable ou de variables indépendantes. La catégorie des GLM englobe la régression linéaire simple, l'analyse de variance, l'analyse de covariance, les mesures répétées, etc.

modèle linéaire généralisé (GLIM), Généralisation plus poussée de l'approche utilisée pour les GLM. Par cette approche unifiée, on estime les paramètres des modèles dans lesquels l'effet suit la loi normale et aussi lorsque les effets appartiennent à n'importe quel membre de la famille exponentielle de distributions, notamment les distributions binomiale, logistique, de Poisson et log-normale. Le toxicologue pourrait utiliser des GLIM pour estimer la dépendance d'un effet quantique ou quantitatif par rapport à une variable indépendante simple telle que la concentration (par la régression), à une structure plus complexe de variables indépendantes telle qu'un traitement de groupe (analyse de variance) ou à des traitements et à des covariables (analyse de covariance). Insuffisamment définie ou circonscrite, la notion de GLIM n'est pas à la portée des non-statisticiens.

monotone, Se dit d'une suite numérique dont chaque élément est soit : *a*) supérieur ou égal à l'élément précédent ; *b*) inférieur ou égal à l'élément précédent.

mortalité partielle (V. *effet partiel.*)

moyenne arithmétique (*moyenne*), Mesure la plus généralement utilisée de la tendance centrale d'un ensemble de données. C'est la somme de toutes les valeurs observées, divisée par le nombre de ces valeurs. Comme, pour la calculer, on tient compte de chaque valeur observée, on peut se représenter la moyenne comme le « centre de gravité » de l'ensemble de données.

moyenne géométrique, Mesure de la tendance centrale d'un ensemble d'observations. Son utilité réside dans le fait qu'elle est moins influencée par les valeurs extrêmes que la moyenne arithmétique, mieux connue. Pour *n* valeurs d'un ensemble, la moyenne géométrique est la racine énième du produit de toutes les valeurs (c'est-à-dire multipliées). On peut aussi la calculer comme l'antilogarithme de la moyenne arithmétique des logarithmes des valeurs.

multicolinéarité (V. *colinéarité.*)

NED (V. *variable normale équivalente.*)

niché (V. *emboîté.*)

niveau de signification, En statistique, probabilité de rejeter l'hypothèse nulle alors qu'elle est vraie. Autrement dit, c'est la probabilité de conclure, à tort, qu'un traitement (par ex. la concentration du toxique) a exercé effet significatif, alors que, en fait, cela est faux. Les toxicologues pourraient aussi utiliser la formule suivante : « ... il existe une différence au niveau de signification de 5 % ». (V. *erreur α* [ou *de première espèce*], *puissance.*)

norme, Niveau défini, exigible, de qualité (V. *ligne directrice relative à la qualité.*)

observation (V. *variable.*)

paramètre, En mathématiques, propriété ou caractéristique d'une *population,* telle que la moyenne ou la médiane. Dans une population donnée, le paramètre possède une valeur constante. Si on tire un échantillon de la population, la moyenne ou la médiane de cet échantillon ne seraient pas des paramètres, mais des *statistiques*. Ces statistiques varieraient presque certainement d'un échantillon à l'autre prélevé dans la même population. Dans les essais de

toxicité, on utilise des échantillons d'organismes, de sorte que les prétendus *paramètres de toxicité* sont des statistiques, considérées comme des estimations de la vraie valeur de la toxicité (le paramètre de la population totale d'organismes). En biostatistique, par convention, des lettres grecques représentent les paramètres de la population et des lettres latines représentent les statistiques des échantillons. Dans l'usage courant, les sens de *paramètre* sont quelque peut incertains. Même l'administration parle de « paramètre » quand il faudrait employer *variable.* Une erreur très répandue consiste à employer le mot « paramètre » dans les listes ou les tableaux de mesures chimiques, pour désigner les substances dosées ; il s'agit en fait de variables, et c'est une erreur qu'il faudrait éviter.

paramètre de toxicité, statistique estimée grâce à un essai. Ce paramètre caractérise les résultats de l'essai (par ex. la *CI p* ou *CL* 50). Il n'est pas recommandé de confondre cette notion et l'effet mesuré sur l'organisme ou la variable observée telle que la taille de l'organisme à la fin de l'essai, bien que cet usage existe (OECD, 2004).

polluant, Substance, matière ou forme d'énergie causant ou capable de causer la *pollution* si elle est rejetée dans l'environnement en quantité suffisante. (V. *pollution* et contaminant.)

pollution, Addition d'une matière ou d'une forme d'énergie telle que la chaleur à un milieu quelconque, en une quantité y causant une altération détectable, qui nuit à une utilisation de ce milieu par les organismes ou par l'homme. Des instances régionales, nationales et internationales ont donné à la pollution des définitions officielles auxquelles il faudrait faire honneur dans les contextes appropriés.

polynôme, Équation d'une *régression multiple* dont certains des termes sont affectés d'un exposant. Par exemple, $\boldsymbol{Y = \alpha + \beta_1 X + \beta_2 X^2 + \beta_3 X^3}$**.**

pondération, Affectation d'un coefficient à certaines valeurs d'une série, en rapport avec leur importance respective, de sorte que ces valeurs exerceront une influence plus grande sur le calcul que l'on effectue. L'objet de la pondération est de compenser certaines irrégularités perçues ou certains manques dans un ensemble de données. On pourrait pondérer une valeur particulière pour montrer qu'il fallait lui accorder plus d'importance parce qu'elle était fondée sur un gros échantillon ou qu'elle était représentative d'un groupe d'observations dont la variance était petite.

population, Collection de toutes les valeurs possibles d'une variable (telle que la longueur de tous les poissons d'un lac). Elle pourrait être constituée de tous les individus du groupe auquel on s'intéresse (tel que les poissons d'un lac). Syn. *univers*. (V. *échantillon.*)

précision, Accord entre les résultats de mesures répétées, souvent évalué par la variance ou l'écart type. Un groupe de mesures pourrait être très précis, mais avoir peu d'*exactitude.* Les mesures pourraient précises et exactes, tout en étant *biaisées.* Si les mesures portent sur plusieurs organismes individuellement (*unités d'échantillonnage*) dans au moins deux enceintes expérimentales (*unités expérimentales*) à une concentration donnée (*traitement*), la variation observée entre les enceintes constitue l'*erreur expérimentale* et elle détermine la précision de la moyenne de mesures à cette concentration. La variation entre les mesures portant sur chaque organisme d'une enceinte constitue l'*erreur d'échantillonnage.* Dans des essais de toxicité quantique, la proportion d'organismes touchés dans une enceinte est l'observation sur l'unité expérimentale, tandis que l'effet présenté par un organisme (touché ou non) est une observation portant sur une unité d'échantillonnage ; la précision est une fonction du nombre d'organismes dans l'enceinte. Dans les méthodes habituelles d'analyse des résultats des essais de toxicité quantique, on réunit les données obtenues sur les répétitions, de sorte que l'on ne peut pas utiliser directement la variation de ces données portant sur chaque concentration.

probabilité (d'un événement), Rapport du nombre de cas favorables à la réalisation de cet événement au nombre total des possibilités.

probit, Unité de mesure de l'écart, par rapport à la moyenne, d'une distribution normale, exprimée en écarts types de la distribution. C'est la *variable normale équivalente*, augmentée d'une valeur constante de 5,0 pour éviter la confusion due aux valeurs négatives d'un côté de la distribution (opération utile à la compréhension, mais en réalité superflue maintenant que les analyses sont confiées à l'ordinateur). L'utilité des probits, dans l'estimation de la CL 50 ou de la CE 50, est de redresser la courbe sigmoïde des probabilités cumulées d'une courbe normale, qui montre le pourcentage d'effet en fonction du logarithme de la concentration. (V. *échelle probit.*)

produit chimique, Tout élément, composé, préparation ou mélange de substances qui pourraient se retrouver associées à un sol, à un sédiment ou à de l'eau ou y être mélangées ou ajoutées.

protocole, Document officiel, exposant ave précision l'ensemble des marches à suivre pendant un essai ou une expérience.

pseudo-répétition, Fausse répétition. Un exemple courant, dans les essais de toxicité, consisterait à qualifier par erreur les organismes d'une enceinte expérimentale de « répétitions ». Dans un test statistique, il s'agit d'une erreur grossière.

puissance, En gros, probabilité de conclure correctement à l'existence d'une différence entre les variables testées. Dans une langue plus soutenue, probabilité de rejeter l'hypothèse nulle alors que, de fait, celle-ci est fausse et devrait être rejetée. En effet, c'est le contraire de la commission d'une erreur *bêta* (ou *de seconde espèce*), qui consiste à ne pas rejeter une hypothèse nulle inexacte. La probabilité de commettre cette erreur est symbolisée par $\boldsymbol{\beta}$, et la puissance est représentée par l'équation $(1 - \boldsymbol{\beta})$. La puissance ne peut pas être fixée de façon directe et précise par l'expérimentateur avant la réalisation de l'essai de toxicité. On peut cependant rendre cet essai plus puissant en augmentant le nombre d'organismes, de répétitions, etc. Le calcul de la puissance à la fin de l'essai est plutôt complexe, mais la puissance est reliée à la *différence significative minimale,* que l'on peut estimer par des méthodes normalisées faisant partie de nombreux tests statistiques portant sur des données quantitatives.

quadratique, Se dit d'un type d'équation pour une régression, qui contient un troisième paramètre et X^2. (Syn. *du second degré*.)

qualité de l'ajustement, Expression ou indice statistique de l'accord des observations avec une distribution théorique ou estimée. La mesure de khi-deux est l'indice habituel, que nous donnerons en exemple. Le khi-deux mesure l'accord entre les fréquences observées et les fréquences théoriques. Le degré d'ajustement (la qualité de l'ajustement) s'exprime par la valeur numérique de khi-deux. [Zar (1999) signale que l'expression « imperfection de l'ajustement » aurait pu être une meilleure désignation, parce que des valeurs de plus en plus grandes de khi-deux signifient un manque de conformité de plus en plus grand entre les observations et la distribution théorique.] Un indice nul signifierait que l'ajustement est parfait, et un indice infini pourrait théoriquement découler d'un ajustement suffisamment mauvais, mais l'indice ne peut pas prendre de valeur négative.

quantique, Se dit d'un *effet* auquel, dans un essai, chaque organisme réagit ou ne réagit pas. Par exemple, un animal pourrait soit mourir, soit survivre ou, encore, se développer normalement ou anormalement. Ce qualificatif s'applique également aux essais et aux données. Par exemple, les données quantiques suivent habituellement une *loi* (*distribution*) *binomiale*. Plus intelligible, le synonyme *dichotomique* est plus fréquent dans les publications de statistique. (V. *binaire, variable binomiale, discret* et *quantitatif.*)

quantitatif, se dit d'un *effet* mesuré pouvant avoir une grandeur exprimée par un nombre entier ou fractionnaire sur une échelle numérique, telle, par ex. la masse ou le poids des organismes à la fin d'un essai. Les données quantitatives obéissent d'habitude à une loi normale. Couramment utilisé par les statisticiens dans le domaine de la toxicologie, en Europe notamment, *continu* pourrait lui être synonyme. *Gradué* était employé dans ce contexte par les premiers

géants de la toxicologie (Gaddum, 1953), mais on ne le considère plus comme approprié. Ce qualificatif s'applique également aux données, aux essais, etc. (V. *quantique.*)

quartile, L'une des trois valeurs qui, dans une série ordonnée de données numériques, divisent la série en quatre parties égales. On divise le nombre d'éléments de la série et non la valeur arithmétique de ces éléments. Le quart de toutes les valeurs de la série seraient inférieures au *premier quartile* et les trois-quarts lui seraient supérieures. Les trois quarts des valeurs seraient inférieures au *3e quartile* et le quart lui seraient supérieures. Le 2e quartile est appelé *médiane,* et la moitié des éléments de la série ordonnée lui sont inférieurs et l'autre moitié supérieurs. (V. *médiane.*) L'*intervalle interquartile* est la valeur absolue de la différence entre le premier et le 3e quartiles. D'habitude, il est assez facile de déterminer les quartiles et la médiane, par examen visuel de la série. Cependant, dans les séries comptant peu d'éléments, le choix des quartiles divisant convenablement la série peut être discutable, et les sources diffèrent sur les définitions et les méthodes précises de calcul. (V. l'annexe R.)

R (*coefficient de corrélation multiple*), Racine carrée du coefficient de détermination (R^2), qui permet d'estimer le coefficient de corrélation multiple (ρ ou *rho*) de la population échantillonnée. *R* égale aussi la corrélation produit-moment de Pearson (d'habitude symbolisée par *r*) entre les valeurs prévues et observées dans une régression. (V. *régression linéaire.*)

R^2 (*coefficient de détermination, coefficient de détermination multiple*), Rapport des sommes des carrés expliqués par un modèle de régression aux sommes totales des carrés par rapport à la moyenne. Dans un contexte de régression, R^2 permet de mesurer la proportion de la variabilité de l'effet mesuré, qu'explique le modèle de régression. (V. *R.*)

régression, Technique statistique déterminant la relation entre deux variables ou plus. La notion englobe l'action et son résultat, la relation après qu'elle a été calculée. La valeur d'une *variable dépendante* (telle que la taille) varie en fonction de la valeur d'une autre ou d'autres variables, la ou les *variables indépendantes* (telles que la concentration). La réciproque n'est pas vraie. La *régression* peut être qualifiée de *simple* si elle ne concerne que deux variables. (V. *régression linéaire, régression non linéaire, corrélation.*)

régression linéaire, méthode statistique d'estimation des paramètres d'un modèle décrivant la relation entre un effet ou une réaction (la *variable dépendante*) et un ensemble de variables explicatives (la ou les *variables indépendantes*). L'adjectif « linéaire » ne s'entend pas de la forme de la courbe, mais de la nature de l'équation décrivant cette dernière. Les modèles linéaires sont relativement simples : on peut en estimer les paramètres (*a, b,* etc.) par résolution d'une seule formule. L'expression « régression linéaire simple » est souvent usitée lorsque l'on utilise seulement une variable explicative. Un modèle linéaire simple serait l'équation bien connue de la droite, $\boldsymbol{Y = a + bX,}$ où *Y* est la variable dépendante, *X* la variable indépendante, $\boldsymbol{a}$ et $\boldsymbol{b}$ les paramètres. Cependant, la régression linéaire peut comprendre des courbes de même que des droites. Par exemple, elle pourrait comprendre un modèle quadratique ($Y = a + bX + cX^2$). Les statisticiens qualifient de « linéaires » les modèles dans lesquels les dérivées partielles du modèle par rapport à un paramètre sont indépendantes de tout autre paramètre. V. *dérivée partielle, régression* et *régression non linéaire.*)

régression multiple, Relation entre une variable dépendante et deux ou plusieurs variables indépendantes. Par exemple $\boldsymbol{Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3}$**.** (V. *polynôme.*)

régression non linéaire, Régression semblable à la *régression linéaire,* mais les *dérivées partielles* d'un paramètre ne sont pas indépendantes d'autres paramètres. L'expression n'a rien à voir avec l'allure de la courbe matérialisant cette relation. La variable dépendante *ne peut pas* être exprimée comme une combinaison linéaire de valeurs des paramètres multipliées par les valeurs de la variable indépendante (SPSS, 1996). La formule de la régression pourrait être multiplicative, par ex. $\boldsymbol{Y = \alpha\beta^X}$, qui est la formule de la croissance exponentielle (Zar, 1999). Pour estimer les paramètres du modèle, il faut procéder par itération. (V. *régression linéaire* et *régression.*)

régression probit (souvent appelée *analyse par la méthode des probits*), Méthode de régression mesurant la relation entre l'intensité d'un stimulus et la proportion des cas chez lesquels un certain effet causé par le stimulus se manifeste (d'après SPSS, 1996). La régression utiliserait normalement la *méthode du maximum de vraisemblance* ou celle des moindres carrés itérativement repondérés, afin d'estimer la CE p et la relation entre le probit de l'effet et le logarithme de la concentration. L'effet analysé est quantique.

régression simple (V. *régression.*)

répéter, V. répétition, à la définition 2.

répétition, **1.** Au sens concret, la répétition d'un *traitement* équivaut à une *unité expérimentale* (V. *bloc*). Enceinte expérimentale renfermant un nombre prescrit d'organismes (= *unités d'échantillonnage*) soit exposés à une concentration (= *traitement*) de matière, soit dans un groupe témoin. Dans un essai de toxicité employant 5 concentrations et 1 témoin, en 3 répétitions (en triple exemplaire), on utiliserait 18 enceintes expérimentales, c'est-à-dire 3 enceintes par traitement. Une répétition doit être une unité expérimentale indépendante, et, par conséquent, la matière se trouvant dans une enceinte ne doit pas avoir de relation avec la matière se trouvant dans une autre enceinte. Tout transfert d'organismes ou de matière d'une répétition à une autre invaliderait l'analyse statistique fondée sur la répétition. L'erreur expérimentale (la variation aléatoire d'une unité expérimentale à l'autre) est estimée à partir des répétitions. (V. *bloc, échantillons réitérés, unité expérimentale, pseudo-répétition, unité d'échantillonnage, traitement.*) — **2.** Action de répéter des ensembles de traitement en groupes, de réitérer un *traitement* ou une *unité expérimentale*.

répétitions de laboratoire (V. *répétition* et *échantillons réitérés.*)

réponse, Synonyme d'*effet.* Ce dernier terme est préféré en toxicologie, parce que les perturbations causées par le toxique chez l'organisme exposé ne sont pas tant une réponse de l'organisme qu'une conséquence de l'action du toxique. Bien que l'expression *dose-réponse* qualifie souvent, de façon générale, les relations dans les essais de toxicité ou les essais biologiques, c'est l'expression « concentration-effet » qui, habituellement, serait plus précise en écotoxicologie. De toute façon, l'effet ou la réponse sont presque toujours la variable dépendante ou *y* dans un modèle statistique.

résidu, Dans une régression, différence entre la valeur observée et la valeur prévue par l'équation.

résistance, capacité, pour un organisme, de retarder la manifestation des effets désignés d'un toxique ou d'un autre agent de l'environnement pendant une période qui dépend de la concentration de l'agent. En fin de compte, l'organisme succombe (d'après Fry, 1947). (Comp. *tolérance.*)

salinité, Traditionnellement, expression de la masse totale de sels dissous dans une masse donnée de solution, en g/kg, en parties pour mille ou en millièmes (‰). Aujourd'hui, on la mesure de façon empirique, à partir des relations standard de densité ou de conductivité et on l'exprime par un nombre pur (APHA *et al.,* 1992).

sédiment, Matériau naturel formé de particules ayant été transportées et s'étant déposées au fond de l'eau. Peut également désigner un substrat artificiel, préparé pour les besoins d'une expérience.

sédiment de référence, Échantillon, prélevé sur le terrain, d'un sédiment que l'on présume non contaminé, possédant des propriétés (par ex. granulométrie, compacité, teneur en matière organique totale) correspondant étroitement à celles de l'échantillon ou des échantillons du sédiment d'essai, sauf le degré de contamination chimique. On le prélève souvent dans un endroit à l'abri de l'influence de la source ou des sources de contamination, mais généralement à proximité des endroits où on prélève le sédiment d'essai. On s'en sert pour décrire les effets de *matrice* dans l'essai

et, aussi, comme témoin et comme diluant pour la préparation de dilutions du sédiment d'essai. (V. *sédiment témoin.*)

sédiment entier, Totalité du sédiment intact ayant subi une manipulation minimale après le prélèvement ou la préparation. Ce n'est pas une forme ni un dérivé du sédiment tel qu'un élutriat ou un sédiment remis en suspension.

sédiment témoin, Sédiment non contaminé, qui pourrait provenir d'un lieu non contaminé ou être préparé (reconstitué). Pour les organismes en culture ou en élevage, ce témoin pourrait être un échantillon de sédiment identique à celui du milieu de culture ou d'élevage. Ce sédiment ne doit être additionné d'aucune autre substance d'essai et il doit permettre un taux acceptable de survie ou des performances acceptables des organismes qui y sont exposés conformément à la méthode. (Comp. *sédiment de référence.*)

série de dilutions, Série de concentrations décroissantes d'essai, dont chacune diffère de la précédente par un coefficient constant (de dilution), par ex. de 100, de 50, de 25, de 12,5 %. On peut obtenir cette série par dilutions successives d'une solution mère d'un déchet ou d'une matière.

seuil (comme dans la CE 50 seuil) [V. *CE 50 initiale.*]

sigma (Σ), Sous la forme majuscule ici représentée, symbole de la sommation d'une suite de termes. La forme minuscule (σ) symbolise le plus souvent l'écart type d'une population.

significatif (V. *signification.*)

signification, Différence intergroupes ou intragroupe, que l'on ne peut pas attribuer uniquement au hasard. Pour ce faire, il faut un test statistique formel. Sauf indication contraire, on suppose un niveau de probabilité de 5 %, c'est-à-dire que la différence ne devrait pas se produire plus de 5 % du temps du seul fait du hasard, si l'expérience ou l'essai étaient répétés de nombreuses fois.

sol de référence, Échantillon prélevé sur le terrain, d'un sol que l'on présume non contaminé, possédant des propriétés (par ex. texture, structure, pH, teneur en matière organique) aussi semblables que possible à celles de l'échantillon ou des échantillons de sol d'essai, sauf qu'il est exempt de la contamination chimique que l'on évalue. On le prélève souvent dans un endroit à l'abri de l'influence de la source ou des sources de contamination, mais généralement à proximité des échantillons d'essai, de sorte qu'il pourrait être soumis à d'autres influences polluantes que celles que l'on étudie. On s'en sert pour décrire les effets de *matrice* dans l'essai et, aussi, comme témoin et comme diluant pour la préparation de dilutions du sol d'essai.

sol entier, Totalité du sol intact ayant subi une manipulation minimale après le prélèvement ou la préparation. Ce n'est pas une forme ni un dérivé du sol tel qu'un élutriat ou un percolat.

sous-échantillon, Subdivision d'un échantillon. Pour le statisticien, les sous-échantillons sont des observations multiples d'un caractère d'une unité expérimentale. Le sous-échantillon doit représenter un prélèvement en un moment unique. Si les prélèvements s'étalent dans le temps, les observations entreraient dans la catégorie des *mesures répétées.*

statistique, Quantité caractérisant une propriété de l'échantillon. (V. *population.*)

sublétal, Nocif pour l'organisme, mais à une concentration inférieure à celle qui est directement mortelle pendant l'essai.

substance, Type particulier de matière, aux propriétés uniformes ; souvent le mot s'applique à un composé chimique.

suite géométrique (Syn. *progression géométrique*), Suite ou progression de nombres dont chaque élément successif est supérieur à celui qui le précède d'un facteur constant (par ex. 3, 6, 12, 24, ...). Les nombres font également partie d'une suite logarithmique.

taux d'erreur expérimentale (V. *erreur.*)

témoin, Se dit, dans une enquête, d'un échantillon reproduisant tous les facteurs qui pourraient influer sur les résultats, sauf la condition ou le traitement particulier à l'étude. Dans un essai de toxicité, le témoin doit reproduire toutes les conditions d'exposition, mais ne pas renfermer de la matière à l'étude (c'est-à-dire aucun toxique). Le témoin sert à vérifier la toxicité apparente due aux conditions de base telles que la qualité d'eau de dilution, l'état de santé des organismes, les effets dus à la manipulation de ces derniers. *Témoin* est synonyme de *témoin négatif.* (V. *témoin positif, témoin de la salinité, témoin du solvant, sédiment témoin, sédiment de référence, sol de référence.*)

témoin de la salinité, Enceinte ou ensemble d'enceintes témoins, séparées, servant, dans un essai de toxicité employant des organismes marins, de *témoin* normal et, aussi, à évaluer tout effet d'une salinité non optimale dans les enceintes expérimentales. L'expression n'aurait pas rapport dans le cas d'essais dans lesquels la salinité de tous les traitements serait ajustée à une valeur optimale standard. Ces essais seraient simplement dotés d'un témoin ayant la même salinité que celle qui existe dans les concentrations expérimentales. C'est le cas des essais effectués en milieu marin en vertu du Programme de suivi des effets sur l'environnement d'Environnement Canada, qui dispose de témoins particuliers, reliés à la technique utilisée pour l'ajustement de la salinité ; voir le § 2.7). Si on n'ajuste pas la salinité dans les concentrations d'essai, on devrait prévoir un témoin à la salinité favorable et, en outre, un ensemble supplémentaire de *témoins de la salinité* reproduisant les salinités expérimentales. On pourrait ainsi isoler l'action nuisible d'une salinité faible (ou forte). Cependant, s'il existait une interaction nuisible entre la salinité anormale et la toxicité de la matière à l'étude, l'ensemble supplémentaire de témoins de la salinité ne révélerait pas cette interaction.

témoin du solvant, Type particulier de témoin qui pourrait être indispensable dans un essai de toxicité, très probablement en milieu aquatique. Il convient à tout essai de toxicité dans lequel on utilise un solvant pour obtenir les concentrations voulues de la substance à laquelle on s'intéresse, si elle est peu soluble. Il faut soumettre à l'essai ce témoin en même temps que le ou les témoins habituels. Le témoin du solvant doit habituellement être soumis aux mêmes conditions que le témoin ordinaire, sauf qu'il doit renfermer la concentration maximale de solvant employée dans l'essai. Pour que les résultats soient satisfaisants, les organismes témoins du solvant doivent se comporter aussi bien que les organismes témoins ordinaires. (V. le § 2.7.2.)

témoin négatif (Syn. *témoin.*)

témoin positif, Essai de toxicité employant un toxique de référence pour évaluer la sensibilité des organismes au moment on évalue une matière et, aussi, la précision des résultats obtenus par le laboratoire sur la substance de référence.

temps létal 50 (V. *TL* 50)

test bilatéral (V. *test unilatéral.*)

test de comparaisons multiples, Méthode statistique permettant de distinguer comment les effets moyens diffèrent statistiquement l'un de l'autre dans une expérience comportant plus de deux traitements. On trouve parfois comme synonyme *test de contraste*. (V. le § 7.5.)

test de contraste (V. *test de comparaisons multiples.*)

test unilatéral, Test statistique s'appliquant à la recherche d'une différence entre une variable et une valeur de référence en se préoccupant du sens de cette différence (par ex. la variable est-elle supérieure à la valeur de référence ?). Dans un *test bilatéral,* on cherche à déterminer si la variable diffère de la valeur de référence sans se préoccuper du sens de la différence, c'est-à-dire la variable est-elle significativement différente ?

TL 50 (*temps létal* 50 %*, temps létal* 50) Durée d'exposition que l'on estime létale pour la moitié des organismes en expérience à une concentration donnée de matière à l'étude. Grâce à des observations successives de la mortalité correspondant à chaque concentration d'une série, on peut estimer le TL 50 de chaque concentration et, parfois, cela a l'avantage de donner une courbe de toxicité plus révélatrice. Les techniques statistiques habituellement utiles à la détermination de la CL 50 ne sont pas valides pour le TL 50.

tolérance, En écotoxicologie, aptitude caractéristique d'un organisme à survivre indéfiniment à des concentrations spécifiées d'un agent de l'environnement. La notion provient d'un travail sur les températures létales pour le poisson, dans lequel on a décrit « la *zone de tolérance* dans laquelle l'animal ne mourra jamais du fait des effets de ce seul agent » (Fry, 1947). [V. *distribution des tolérances, résistance.*]

toxicité, Capacité propre d'une matière de provoquer des effets nocifs chez les organismes vivants.

toxicité aiguë, Effet négatif (létal ou sublétal), discernable, provoqué chez l'organisme en expérience dans une courte période d'exposition à une matière, habituellement de quelques jours quand il s'agit de gros organismes.

toxicité chronique, Manifestation des effets négatifs d'un poison reliés à des modifications des processus fondamentaux tels que ceux du métabolisme, de la croissance ou de la reproduction. On pourrait cependant évaluer l'effet chronique par la mortalité ou par la longévité.

toxicologie, Au sens large, science qui précise les limites de sécurité de l'emploi d'agents chimiques. Ses études doivent être conçues dans ce but. C'est pourquoi, rien ne limite leur échelle (moléculaire ou écosystémique), la gamme des disciplines scientifiques auxquelles elles peuvent faire appel ni l'endroit où se trouvent leurs outils (au laboratoire ou sur le terrain). [V. *écotoxicologie.*]

toxine, Substance toxique, protéine particulièrement, élaborée par des cellules ou des organismes et capable de provoquer la maladie ou d'autres effets nocifs quand elle est introduite dans un organisme. La toxine peut également stimuler la production d'une antitoxine. Un exemple serait la saxitoxine, toxine paralysante des algues, produite par les dinoflagellés marins (phénomène des « eaux rouges »). Les médias et des écolos négligents ont galvaudé ce mot en l'assimilant à toutes sortes de toxiques.

toxique, adj. Se dit d'une substance, d'une matière ou d'un produit présent en quantité suffisante pour causer des effets nocifs ou pouvant être nocifs pour les organismes vivants. N. m. Substance, produit ou matière pouvant provoquer des effets nocifs chez les organismes vivants. Poison.

toxique de référence, Substance étalon servant à mesurer la sensibilité des organismes en expérience et à aider à établir la validité des données toxicologiques obtenues sur la matière à l'étude. Dans la plupart des cas, on réalise un essai de toxicité avec un toxique de référence pour évaluer : *a*) la sensibilité des organismes au moment de l'essai sur la matière ; *b*) la précision des résultats obtenus par le laboratoire, au cours d'une période pendant laquelle a été effectué un nombre plus ou moins grand d'essais avec ce toxique de référence.

traitement, De façon générale, application d'un facteur dont on veut mesurer l'effet sur une *unité d'échantillonnage.* Au sens strict, dans les essais de toxicité, c'est un facteur appliqué aux organismes en expérience, afin de mesurer son effet sur ces derniers. Habituellement, il s'agirait d'une concentration de matière potentiellement toxique. Le

traitement pourrait englober plusieurs récipients exposés à la même concentration, chacun d'entre eux constituant une *unité expérimentale* et, également, une *répétition.* Dans les essais portant sur un sédiment ou un sol, la matière précise que l'on soumet à l'essai (par ex. sédiment ou sol d'un emplacement, sol de référence ou sol témoin négatif) provenant d'une station particulière d'échantillonnage. (V. *unité d'échantillonnage.*)

transformation arc sinus, Transformation applicable aux données constituées de proportions ou de pourcentages, qui tendent à former une distribution binomiale. Il s'agit, par cette transformation, de rendre les variances homogènes et la distribution presque normale, afin d'effectuer des analyses statistiques paramétriques. La transformation est l'arc sinus de la proportion en question. La notation abrégée est *arc sin*. De nombreux logiciels et calculettes scientifiques permettent de l'effectuer. On peut aussi se servir d'une table, consultable dans la plupart des manuels de statistique. Cette transformation, dont on pourrait bien se passer aujourd'hui, était utile avant l'avènement des moyens modernes de calcul qui épargnent la corvée des calculs manuels.

unité d'échantillonnage, Unité d'observation dans une *unité expérimentale.* Par exemple un organisme dans un récipient renfermant des organismes exposés à un *traitement* donné. (V. *répétition.*)

unité de toxicité (UT), Expression du pouvoir toxique d'un déchet ou d'une substance se trouvant dans un milieu tel qu'un sol, un sédiment, l'eau ou l'air. Ce pouvoir s'exprime en multiples (ou fractions) d'un paramètre standard de toxicité. Le nombre d'unités de toxicité d'un déchet tel qu'un effluent se calculerait comme suit : 100 % (c'est-à-dire l'effluent non dilué) divisé par la concentration du paramètre de toxicité exprimé en pourcentage (par ex. un effluent dont la CL 50 serait de 10 % aurait 100/10 = 10 UT létales). Dans le cas d'un toxique se trouvant dans un substrat ou un milieu, on peut prendre comme exemple une substance chimique dissoute dans l'eau. On calculerait son *pouvoir toxique* (en *unités toxiques létales*) par le quotient de sa concentration réelle dans l'eau divisée par sa CL 50. Pour le calcul du *pouvoir toxique* en *unités toxiques sublétales,* le quotient aurait comme dénominateur un paramètre défini de toxicité sublétale (tel que la CI 25). Par exemple, si la substance était présente dans l'eau à la concentration de 5 mg/L et que sa CI 25 était de 10 mg/L, on aurait 5/10 = 0,5 UT sublétales, c'est-à-dire la moitié de la concentration à laquelle s'exerce l'effet sublétal. L'unité de toxicité est un nombre pur n'exprimant pas de concentration de substance. C'est un concept pratique, puisque le nombre d'unités de toxicité augmente proportionnellement au pouvoir toxique.

unité expérimentale, Plus petite unité ou élément indépendant dans un essai de toxicité, auquel on applique un traitement. L'unité expérimentale présente un effet que l'on mesure et qui devient une donnée. Un exemple serait une enceinte renfermant des organismes dans un essai de toxicité. (Les organismes de cette enceinte seraient les *unités d'échantillonnage.*) Si au moins deux enceintes étaient exposées à un traitement, chaque enceinte constituerait à la fois une unité expérimentale et une *répétition.* (V. *unité d'échantillonnage* et *bloc.*)

UT (V. *unité de toxicité.*)

variable, Caractéristique dont la valeur diffère d'un individu, d'un cas ou d'une observation à l'autre. Ainsi, elle caractérise les individus ou les cas dans une *population* d'individus ou de cas. Ce pourrait être la concentration d'une substance, la hauteur de plantes, le nombre de petits ou des éléments semblables. La valeur mesurée ou consignée de la variable constitue une *observation.* La variable peut être *continue,* prenant n'importe quelle valeur dans un intervalle possible (tel que la concentration d'une substance ou le poids d'une larve de chironomidé. Elle peut, d'autre part, être *discrète,* ce qui signifie qu'elle peut prendre toute valeur positive ou négative telle que 0, 1, 2, 3, par ex. le nombre de feuilles d'une plante. Les deux désignations correspondent, respectivement, aux données *quantitatives* et *quantiques.* Dans une analyse, une *variable indépendante* serait celle qui est fixée, habituellement par l'expérimentateur, qui s'en sert pour prévoir la valeur correspondante de la *variable dépendante.* La valeur de cette dernière est déterminée par le choix de la variable indépendante. Dans un essai de toxicité, les concentrations seraient la variable indépendante, et l'effet serait la variable dépendante. (V. *variable binomiale, paramètre.*)

variable aléatoire binomiale, V. *variable binomiale.*

variable binomiale ou *variable aléatoire binomiale,* Nombre d'individus possédant, dans une expérience, l'une des deux caractéristiques quantiques ou binaires possibles (par ex. la mort).

variable dépendante (V. *variable.*)

variable discrète (V. *variable.*)

variable indépendante (V. *variable.*)

variable normale équivalente (*NED,* pour *normal equivalent deviate*), Écart type d'une distribution normale, associé à une probabilité particulière. Autrement dit, c'est une unité de dispersion d'une distribution normale, exprimée par l'écart type de cette distribution. 1 *NED* correspond à l'intervalle de ± 1 écart type par rapport à la moyenne. Le probit est tout simplement un *NED* auquel on a ajouté 5 pour éviter l'obtention de valeurs négatives pour un côté de la distribution.

variance, Caractéristique de la dispersion des observations individuelles sur un échantillon par rapport à la valeur moyenne de cet échantillon. On la calcule comme suit : *a*) on soustrait la moyenne de chaque observation ; *b*) on élève au carré chacune de ces différences ; *c*) on somme ces carrés ; *d*) on divise cette somme par le nombre d'observations moins un. Le symbole de la variance est s^2. (La variance de la population théorique d'où on a tiré l'échantillon aurait σ^2 comme symbole, et on l'estimerait à partir d'un échantillon par la méthode que nous venons de décrire, sauf que, à l'étape *d*), on diviserait le résultat de la sommation par le nombre d'observations.) On omet généralement les unités de la variance ; ce sont les carrés des unités utilisées à l'origine, ce qui pourrait ne pas avoir du sens. (V. *écart type.*)

variante expérimentale (V. *traitement.*)

Remerciements

Le présent document a été rédigé par John B. Sprague (Sprague Associates Ltd., de Salt Spring Island [C.-B.]) avec l'apport technique direct de Barry A. Zajdlik (Zajdlik & Associates Inc., de Rockwood [Ont.]) et d'après les propositions de Glenn F. Atkinson (Atkinson Statistical, Calgary). Il s'inspire de guides, de rapports et de publications sur la toxicologie et la statistique ainsi que des idées de scientifiques et de techniciens d'administrations publiques, de l'industrie et d'universités d'un peu partout au Canada et ailleurs. Richard P. Scroggins (chef de la Division des méthodes biologiques, Environnement Canada [EC], Ottawa) a été le responsable scientifique du projet, dont il a accompagné la réalisation par son aide technique et ses conseils. Stella Wheatley (Polaris Scientific and Technical Editing, Ottawa) a corrigé et mis en forme le document et préparé certaines figures. Juliska Princz (EC, Ottawa) a rassemblé et uniformisé la description des modèles du § 6.5.8 et les consignes d'emploi de SYSTAT de l'annexe O, d'après des méthodes publiées récemment par Environnement Canada.

Nous sommes reconnaissants de l'aide inestimable que nous ont offerte les personnes suivantes, sous forme d'observations et de conseils sur une ou plusieurs ébauches du document : Larry W. Ausley (min. de l'Environnement de la Caroline du Nord, Raleigh) ; Uwe Borgmann (Institut national de recherche sur les eaux [INRE], Burlington [Ont.]) ; Kenneth G. Doe (EC, Moncton) ; Natalie Feisthauer (Stantec Consulting, Guelph, [Ont.]) ; Hector F. Galicia (Springborn Smithers Laboratories (Europe) AG, Horn, Suisse) ; John W. Green (DuPont, Newark, Delaware) ; Christine S. Hartless (USEPA, Washington) ; Janet McCann (U. de Waterloo, Ont.) ; Donald J. McLeay, (McLeay Environmental Ltd., Victoria) ; Cathy McPherson (EVS Environment Consultants, North Vancouver [C.-B.]) ; Jennifer Miller (Miller Environmental Sciences Inc., Innisfil [Ont.]) ; Mary Moody (Saskatchewan Research Council, Saskatoon) ; Serge Morissette (min. de l'Environnement, Sainte-Foy [Qc]) ; Marion Nipper (U. A & M, Corpus Christi, Texas) ; Niels Nyholm (U. technique du Danemark, Lyngby) ; R. Jeanette O'Hara Hines (U. de Waterloo) ; Juliska Princz (EC, Ottawa) ; Hans Toni Ratte (École technique supérieure de Rhénanie-Westphalie, Aix-la-Chapelle) ; Jim Reid (ESG International, Guelph) ; Julie E. Schroeder (min. de l'Environnement et de l'Énergie de l'Ontario [MEEO], Etobicoke) ; Wout Slob (Institut national pour la santé publique et l'environnement, Bilthoven, Pays-Bas).

Nous remercions les personnes suivantes qui ont fourni une aide importante sous forme d'analyses informatiques, de renseignements, de rapports et d'autres moyens tangibles ou qui ont organisé des ateliers sur le projet, y ont participé ou en ont résumé les délibérations : Howard Bailey (EVS Environment Consultants, North Vancouver) ; Joy Bruno (EC, North Vancouver) ; Craig Buday (EC, North Vancouver) ; Curtis Eickhoff (BC Research Inc., Vancouver) ; Paula Jackman (EC, Moncton) ; Nicky Koper (U. de l'Alberta à Edmonton) ; Nancy Kruper (EC, Edmonton) ; Don Larson (IRC Consultants, Richmond [C.-B.]) ; Michelle Linssen (EC, North Vancouver) ; Tim Moran (Pollutech Enviroquatics, Point Edward [Ont.]) ; David Moul (EC, North Vancouver) ; Michael D. Paine (Paine, Ledge and Associates, North Vancouver) ; Janet Pickard (BC Research Inc., Vancouver) ; Linda Porebski (EC, Gatineau [Qc]) ; Danielle Rodrigue (EC, Ottawa) ; Gladys L. Stephenson (ESG International Inc., Guelph) ; Armando Tang (EVS Environmental Consultants, North Vancouver) ; Becky-Jo Unis (Hydroqual Laboratories, Calgary) ; Graham van Aggelen (EC, North Vancouver).

Introduction

Les essais de toxicité sont de puissants outils d'étude et de résolution des problèmes de contamination et de pollution de l'environnement. Cependant, il faut analyser convenablement leurs résultats pour obtenir des estimations valables des paramètres de toxicité. Le présent document vise à faciliter le choix des bonnes méthodes d'analyse statistique des résultats des essais employés en écotoxicologie. Il est particulièrement destiné à servir de concert avec plus de 20 méthodes d'essai de toxicité publiées par Environnement Canada, dans lesquelles on emploie des microorganismes, des plantes aquatiques et terrestres, des invertébrés et des poissons (EC, 1990a-c ; 1992a-f ; 1997a, b ; 1998a, b ; 1999b ; 2000a, b ; 2001a ; 2002a ; 2004a, b et 2007 ; v. l'énumération de l'annexe A). Le document se concentre sur les méthodes mathématiques et statistiques d'analyse des résultats ; un autre guide traite des méthodes générales et des interprétations concernant l'écotoxicologie (EC, 1999a).

1.1 Buts et objectifs du document

Repères

- *Le principal objectif du document est d'aider à établir de bonnes pratiques statistiques dans les laboratoires canadiens où l'on effectue des essais de toxicité relevant de programmes d'Environnement Canada.*
- *Les tests statistiques en usage font l'objet d'une discussion, au terme de laquelle on précise les méthodes à privilégier et celles qui sont prometteuses. Des exemples sont présentés.*
- *Les explications s'adressent principalement au nouveau personnel de laboratoire. On insiste sur la mise à l'épreuve des normes plutôt que sur les projets de recherche.*
- *Des conseils sont donnés pour reconnaître les types « difficiles » de données et pour s'en occuper. On explique des erreurs répandues.*

Le présent document vise simplement à donner des renseignements dans trois domaines :

a) Des conseils supplémentaires aux utilisateurs des essais monospécifiques d'Environnement Canada. Ces conseils s'adressent principalement au nouveau personnel de laboratoire, plutôt qu'aux expérimentateurs chevronnés.

b) Des explications de la raison d'être statistique des méthodes employées dans les essais de toxicité. Cependant, le document n'est pas une introduction à la statistique.

c) Des observations sur les tests statistiques existants et sur certaines approches profitables qui pourraient devenir accessibles.

Les objectifs fondamentaux du document ont été définis par un comité consultatif de la statistique et d'autres intéressés qui se sont réunis après le 20e Atelier annuel sur la toxicité en milieu aquatique (*Annual Aquatic Toxicity Workshop*), qui a eu lieu à Québec en 1993 (Miller *et al.,* 1993). Ces objectifs sont les suivants :

(1) Donner des conseils sur les méthodes statistiques applicables aux essais biologiques, favorisant ainsi la normalisation des méthodes de calcul des paramètres expérimentaux de toxicité ;

(2) Fournir des renseignements de base sur les caractéristiques, les points forts et les points faibles des diverses méthodes statistiques et sur l'importance de leurs hypothèses ;

(3) Fournir des méthodes d'évaluation pour déterminer si les résultats d'une expérience répondent de façon définitive aux questions initialement posées ;

(4) Donner des exemples d'application des méthodes statistiques et d'interprétation de leurs résultats ;

(5) Donner des conseils sur la reconnaissance des données « difficiles » et la façon de se tirer

d'affaire avec elles.

À Québec, on a présenté des documents d'information, et la discussion de neuf sujets a abouti à la formulation de recommandations précises, auxquelles il est donné suite dans les parties appropriées du présent document. Il est donné suite aux objectifs 3 et 4, comme il se doit, dans l'ensemble du document, lorsque l'on décrit des méthodes particulières.

On a augmenté le glossaire en l'enrichissant d'exemples et d'explications permettant aux expérimentateurs d'appliquer leurs connaissances générales aux particularités des essais d'écotoxicité.

Des exemples agrémentent l'information donnée dans le document, notamment la mention des pièges et des lacunes. Les expérimentateurs chevronnés pourraient avoir l'impression que l'on insiste trop sur les erreurs communes, mais il importe qu'il en soit ainsi. Des données prélevées à la faveur de programmes réglementaires nationaux ont révélé, dans les modes opératoires, des erreurs relativement mineures mais fréquentes. Cela résulte peut-être d'une pénurie de personnel de laboratoire expérimenté ou qui s'y connaît. Pourquoi la prochaine cohorte devrait-elle répéter des erreurs banales ?

En conséquence, le document s'attache en grande partie aux méthodes courantes et établies d'écotoxicologie et il ne tente pas d'innover. Bien qu'il attire l'attention sur les techniques nouvelles et avancées qui semblent être à la fine pointe, il ne peut pas s'étendre sur les méthodes d'analyse en développement. Certaines des méthodes nouvelles s'imposeront sans doute en raison des avantages qu'elles offrent, tandis que les autres retomberont dans l'oubli. En général, ces méthodes nouvelles sont proposées par les spécialistes des équipes statistiquement branchées. De même, il semble que les études canadiennes de problèmes locaux complexes de toxicité et de pollution bénéficient généralement et directement des conseils de spécialistes en matière de statistique dont elles ont besoin.

Le document ne vise pas à donner des conseils sur les programmes de recherche fondamentale. Les chercheurs et ceux qui appliquent des techniques statistiques avancées pourraient trouver des conseils utiles dans les sections 4 et 6 du présent document, dans un document d'une organisation internationale (OECD, 2004) et chez les auteurs cités dans les paragraphes qui suivent.

On peut trouver des conseils généraux en statistique dans des manuels tels que ceux de Snedecor et Cochran (1980), Steel *et al.* (1997), Zar (1999) et Wardlaw (1985), ce dernier étant rédigé dans un style qui plaît particulièrement aux biologistes. On peut accéder à la base statistique de la toxicologie et d'autres études de l'environnement grâce à Newman (1995), Gad (1999), Manly (2000) et Millard et Neerchal (2000). Finney (1971 ; 1978), Ashton (1972), Hewlett et Plackett (1979) ou Hubert (1992) décrivent des sujets reliés à la toxicologie classique et à d'autres sujets plus spécialisés. Collett (1991) et Fleiss (1981) donnent des conseils sur l'analyse des proportions et des données binaires, le fondement des effets quantiques. Un livre anecdotique intitulé *The Lady Tasting Tea* (Salsburg, 2001) procure une initiation philosophique générale aux grandes idées de la statistique appliquée. D'autres précis, aux titres apparemment pertinents pourraient ne pas être d'une aide immédiate pour l'analyse statistique des essais (OECD, 1995 ; Grothe *et al.,* 1996). Enfin, on trouve des sources prodigieuses de renseignements (et, parfois, de désinformation) dans Internet, aux sites sans cesse changeants. Certains sites sont utiles, pour ce qui est de communiquer les notions générales ou des techniques statistiques particulières. Certains sites excellents équivalent à des chapitres de manuels ou à des notes de cours.

1.2 Mode d'emploi du document

Pour réaliser un essai de toxicité, on commencerait d'habitude par une méthode publiée par Environnement Canada ou une autre organisation. Chacun de ces documents précise les méthodes statistiques à utiliser, en y allant de descriptions qui conviennent à la plupart des objectifs.

Pour obtenir plus d'explications, l'expérimentateur pourrait consulter l'une des sections 3 à 8 du présent document, qui portent sur différents types de tests et d'analyses. Ces sections débutent au niveau d'une méthode individuelle et donnent des conseils supplémentaires sur l'analyse, la façon d'éviter les pièges, en motivant le choix des méthodes. Les modes opératoires particuliers d'essai décrits dans chaque méthode publiée par Environnement Canada sont

définitifs et on devraient s'y conformer dans les programmes de ce ministère. Le présent document n'annule et ne remplace aucune méthode particulière d'essai.

Les novices pourraient souhaiter lire rapidement les sections 2, 9 et 10. La section 2 traite d'aspects des plans d'expérience, de certaines erreurs fréquentes et elle offre des renseignements de base. La section 9 est plus générale, elle donne des renseignements statistiques de base et elle traite de méthodes permettant les tests de différences. Dans la section 10, on donne des conseils sur certains types difficiles de résultats, notamment les observations aberrantes et l'effet de stimulation à de faibles concentrations de toxique. Les lecteurs peuvent consulter la rubrique « Repères » au début des sous-sections, pour décider des parties qui pourraient leur être utiles. Comme nous l'avons mentionné, le glossaire est détaillé de façon à offrir des conseils supplémentaires.

Les annexes constituent la deuxième partie du document. Elles renferment, au profit des lecteurs intéressés, des renseignements techniques ou détaillés étayant les affirmations du corps du texte.

1.3 Principales catégories d'essais

Repères

- *Il existe deux grandes catégories d'essais :* a*) les essais à concentration unique visant à comparer une matière à une matière témoin ou à une matière de référence ;* b*) les essais à concentrations multiples, visant à estimer une CE p, une CI p ou une CSEO.*
- *Les essais à concentrations multiples peuvent se subdiviser en :* a*) essais à variables quantiques (chaque organisme réagit ou ne réagit pas) ;* b*) les essais à variables quantitatives ou continues (qui s'intéressent, par exemple, au poids des individus).*
- *Les essais de mesure d'un double effet portent souvent sur des mesures quantiques et quantitatives, qui, pour le moment, sont le mieux analysées séparément.*

Une première division entre les types d'essais permet de séparer ceux qui utilisent une seule concentration de la matière à l'étude et ceux qui en utilisent plusieurs.

Les essais à concentration unique comparent les effets toxiques d'un échantillon à ceux d'une matière témoin (ou d'une matière de référence ou d'un autre échantillon, emplacement ou condition particulière). Par exemple, on pourrait comparer un seul échantillon de sédiment à un sédiment de référence. Ces essais à concentration unique occupent la partie gauche de la fig. 1. Des variantes pourraient comprendre l'emploi d'un seul échantillon ou d'un certain nombre d'échantillons provenant de différents emplacements, comparés simultanément à une matière témoin ou à une matière de référence. Il pourrait ou ne pourrait pas y avoir répétition. Ces types d'essais sont décrits dans la section 3.

Les essais à plusieurs concentrations utilisent plusieurs concentrations fixes et un témoin, afin d'estimer une CE p, une CI p ou une CSEO. Ces essais occupent les parties médiane et droite de la fig. 1, qui donne une vue d'ensemble générale des types d'essais.

Ces deux types d'essais permettraient l'observation d'un effet quantique ou d'un effet quantitatif (fig. 1). Dans les essais quantiques, le dénombrement direct des organismes exposés permet de les classer dans le groupe des organismes non touchés ou touchés, c'est-à-dire que les données sont *binaires* ou dichotomiques. Les résultats sont le mieux ajustés à une distribution binomiale et sont analysés par des techniques statistiques convenant à une telle distribution (par ex. le test du khi-deux). Cependant, la plupart des essais quantiques en écotoxicologie sont des essais de toxicité létale. L'analyse emploie habituellement la régression probit ou logit ou une méthode de remplacement. Le paramètre habituel de toxicité estimé est la *concentration létale médiane* (CL 50) ou la *concentration efficace 50* (CE 50), expression plus générale qui englobe les effets sublétaux. Les essais quantiques sont décrits dans la section 4.

Les essais quantitatifs mesurent un effet variant de façon continue tel que la taille d'un organisme. Ces essais ont déjà été qualifiés de « gradués » (Gaddum, 1953), qualificatif qui ne semble plus convenir. Les

résultats de ces essais peuvent être qualifiés de données « continues ». Le paramètre habituel de toxicité est la *concentration inhibitrice pour un pourcentage donné d'inhibition des performances des organismes* (CI p). Par exemple, la CI 25 pourrait représenter la concentration correspondant à une diminution de 25 % du poids des organismes par rapport au poids des organismes des groupes témoins. Les résultats obéissent souvent à la loi normale bien connue (de fait, une courbe normale cumulée). Idéalement, on estimerait les paramètres de toxicité par régression, et les méthodes de régression ont récemment été prescrites dans certains essais normalisés publiés par Environnement Canada (§ 6.5.8). Une autre méthode, moins privilégiée cependant, consiste à utiliser un test d'hypothèse pour évaluer la CSEO et la CEMO, comme il est décrit dans la section 7.

Certains essais aboutissent à des effets doubles et ils comportent habituellement un effet quantique tel que la mortalité et un effet quantitatif, comme sur le poids ou la reproduction. Pour le moment, les résultats des essais de mesure d'un double effet devraient être analysés séparément (partie centrale de l'organigramme de la fig. 1), faute d'une technique qui aurait été mise au point pour analyser les effets corrélés. On adopte habituellement comme paramètre de toxicité estimé grâce à l'essai la concentration efficace inférieure. Les essais de mesure d'un double effet sont décrits dans la section 8.

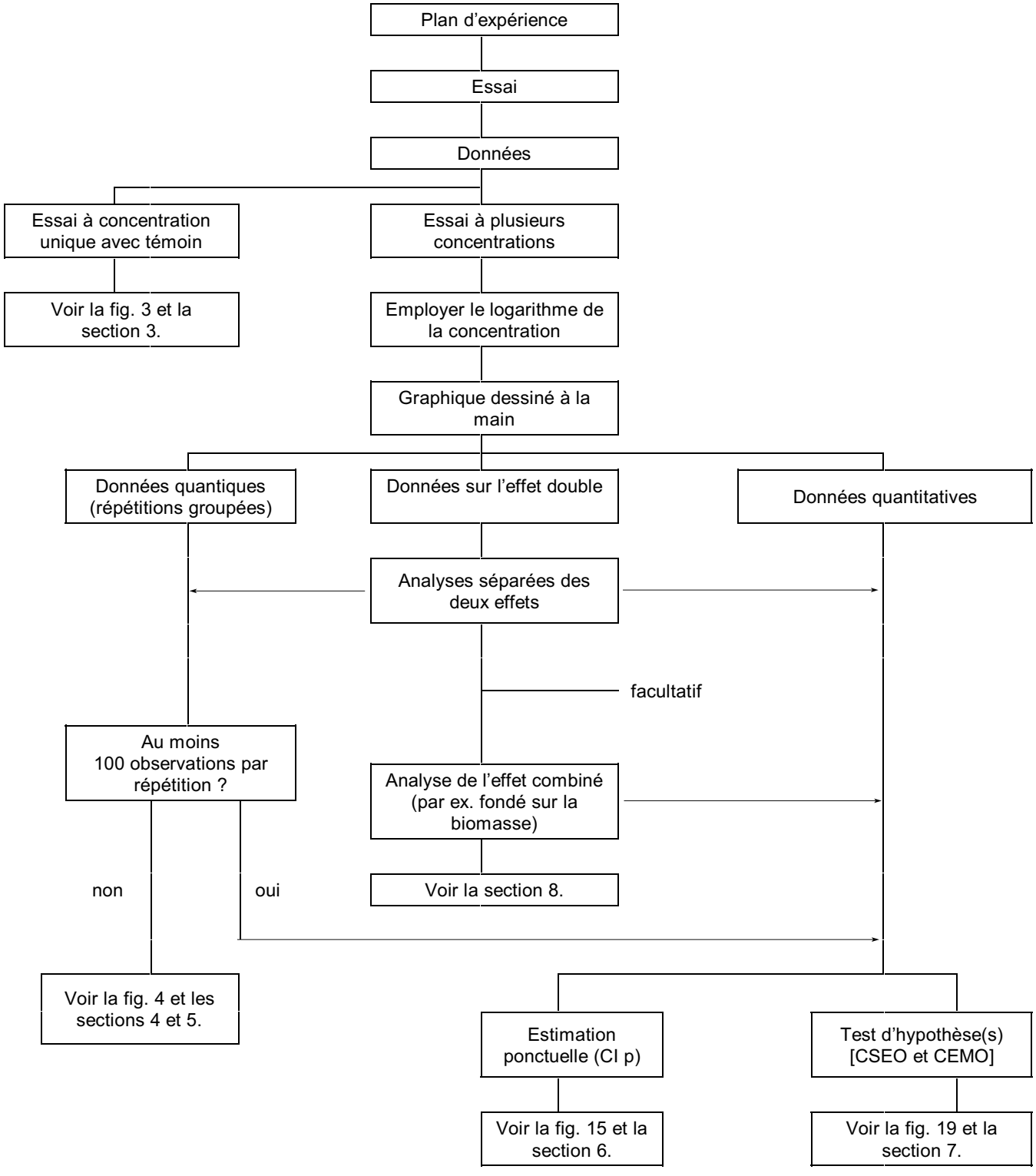

Plan d'expérience
Essai
Données
Essai à concentration unique avec témoin
Essai à plusieurs concentrations
Voir la fig. 3 et la section 3.
Employer le logarithme de la concentration
Graphique dessiné à la main
Données quantiques (répétitions groupées)
Données sur l'effet double
Données quantitatives
Analyses séparées des deux effets
facultatif
Au moins 100 observations par répétition ?
Analyse de l'effet combiné (par ex. fondé sur la biomasse)
non
oui
Voir la section 8.
Voir la fig. 4 et les sections 4 et 5.
Estimation ponctuelle (CI p)
Test d'hypothèse(s) [CSEO et CEMO]
Voir la fig. 15 et la section 6.
Voir la fig. 19 et la section 7.

Figure 1. — Organigramme des principales catégories d'essais d'écotoxicité traitées dans le présent document.

Planification générale et analyse

Les meilleures méthodes statistiques ne sauraient remédier à une expérience mal planifiée. Les méthodes d'essai biologique publiées par Environnement Canada comprennent des conseils sur la planification des expériences, qu'il faudrait suivre pour obtenir des données qui se prêtent à l'analyse statistique. En particulier, l'expérimentateur ne devrait jamais faire fi des instructions sur la randomisation, les répétitions ou les témoins (v. les § 2.4, 2.5 et 2.7). Les § 2.1, 2.2 et 2.3 traitent d'autres aspects importants des plans d'expérience.

La variabilité des résultats peut parfois être réduite grâce à l'utilisation judicieuse d'un nombre d'organismes, de répétitions ou de concentrations supérieur au nombre minimal requis. Ce type d'amélioration serait particulièrement souhaitable dans les essais qui serviront à l'homologation (enregistrement) de nouveaux produits chimiques ou à des poursuites judiciaires.

2.1 Participation d'un statisticien

Qu'un statisticien doive intervenir à toutes les étapes d'un essai, notamment dans la planification, l'analyse et l'exposé des résultats, est un truisme. Souvent, cela est plus facile à dire qu'à faire, particulièrement dans les petits laboratoires ou les essais confiés au privé ; cependant, le principe n'en est pas moins vrai. En cette époque de communications modernes, on devrait pouvoir mettre au point un système de consultation rapide et économique. Le remède pourrait se trouver dans l'organisation, de temps à autre, de grandes séances de prestation de conseils sur des approches et des solutions souhaitables à d'éventuelles situations. Le présent document ambitionne de guider l'expérimentateur, afin que sa consultation d'un statisticien soit plus fructueuse.

Parfois, le conseil que donnerait le statisticien pourrait consister simplement à réaliser un essai de toxicité ordinaire. On devrait s'entendre sur le plan d'expérience et les méthodes d'analyse, et le statisticien pourrait avertir l'expérimentateur des difficultés potentielles. Si on a des motifs de craindre une distribution irrégulière des données, il faudrait envisager cette éventualité à l'étape de la planification de l'essai. Si un effet est susceptible de se manifester chez les organismes témoins, il faut également en tenir compte pendant la planification.

En même temps, l'expérimentateur, qui pourrait être biologiste, devrait ne pas oublier ses priorités et tenter de parvenir à un juste équilibre entre ces dernières et les avis du statisticien. Si on lui demande de restreindre les limites d'erreur, le statisticien peut à peine se retenir de recommander plus de concentrations, plus d'organismes, etc. L'expérimentateur doit prendre des décisions difficiles, pour parvenir à un juste équilibre entre ces recommandations et les questions pratiques de coûts, de délais, d'installations et de priorités de travail. (À celui qui a la bosse des mathématiques, les statistiques bayésiennes permettent de combiner les notions de probabilités et de coûts pour déterminer s'il est rentable ou non d'obtenir des renseignements supplémentaires. On peut trouver une introduction à ce sujet dans Morissette [2002] ; elle porte principalement sur le prélèvement d'échantillons de sédiments contaminés, mais elle s'applique à d'autres sujets.)

2.2 Sélection des concentrations

Repères

- *Le choix des concentrations est un aspect important et difficile du plan d'expérience. Si l'on pouvait prévoir les résultats, le plan idéal serait de posséder plusieurs concentrations auxquelles correspondent un intervalle médian d'effets, et d'autres concentrations réparties également au-dessus et au-dessous et auxquelles correspondraient des effets négligeables à grands.*
- *Le problème le plus fréquent dans le choix des concentrations est que ces dernières sont trop rapprochées les unes des autres. Toutes pourraient se révéler soit trop élevées, soit trop*

faibles, ce qui gâcherait l'essai.

- *Le choix de concentrations séparées par des distances importantes fait courir le risque de ne pas pouvoir observer d'effets « partiels » ou des effets de milieu de gamme, mais ce risque est moins grave.*
- *Un moyen utile d'éviter ce problème serait d'effectuer une recherche préliminaire de la gamme de concentrations à utiliser et/ou d'utiliser un nombre relativement élevé de concentrations dans l'essai définitif.*
- *Quel que soit le choix de concentrations, une bonne planification exige l'emploi d'une suite géométrique régulière.*

Dans un essai à plusieurs concentrations, le choix des concentrations appropriées est l'aspect le plus important du plan d'expérience. Un mauvais choix est la cause la plus fréquence d'obtention de résultats « difficiles ». L'erreur la plus répandue consiste à choisir des concentrations trop rapprochées les unes des autres. L'expérimentateur pourrait conjecturer (mais à tort) la valeur prévue du paramètre de toxicité et faire un choix mauvais et malheureux de concentrations d'essai. Ce choix pourrait le conduire à la pire des situations possibles, c'est-à-dire la manifestation d'effets graves à toutes les concentrations ou, au contraire, d'effets faibles. L'essai serait un échec, particulièrement s'il avait utilisé un échantillon prélevé sur le terrain que l'on ne pourrait pas reproduire. On donne des exemples dans le § 10.4.

Pour déterminer une CE p ou une CI p, on devrait toujours s'efforcer de disposer de concentrations qui lui sont à la fois supérieures et inférieures. Faute de telles concentrations, il est toujours déconseillé d'estimer cette valeur par extrapolation, et cela est souvent impossible. Malheureusement, on ne pourrait choisir parfaitement les concentrations que si l'on connaissait d'avance les résultats de l'essai : l'expérimentateur est donc obligé d'utiliser son jugement. Ce dernier peut être considérablement amélioré s'il effectue un essai préliminaire de recherche de la gamme de concentrations à utiliser, même si cet essai est assez approximatif. Ensuite, il peut améliorer considérablement le plan d'expérience de l'essai de toxicité en augmentant le nombre de concentrations à utiliser et en les espaçant suffisamment.

Quels que soient l'objet et le plan général d'expérience de l'essai, il importe de choisir une suite géométrique régulière de concentrations. Chaque concentration doit présenter une augmentation constante par rapport à celle qui la précède. Par exemple, le doublement des concentrations pourrait donner la suite suivante : 4, 8, 16, 32, 64, etc. À première vue, l'écart paraît grand entre 32 et 64, mais, pour l'organisme exposé l'augmentation représente exactement le même doublement de la stimulation d'origine chimique qu'entre 4 et 8. Elle représente exactement le même doublement que celui qui se placerait dans une suite de concentrations 10 fois plus faibles, c'est-à-dire entre 3,2 et 6,4. Quelle que soit la région de la suite qui se révèle être l'intervalle important, la même proportionnalité existe. Cela aide à équilibrer la répartition des résultats, et le choix est fondamental pour tous les calculs ultérieurs (v. le § 2.3 pour plus de détails).

2.2.1 Influences contraires

La précision et les limites de confiance d'un paramètre de toxicité dépendent en tout ou partie des facteurs suivants :

a) Le nombre de concentrations auxquelles correspondent des effets « médians » ou partiels ;
b) La dispersion des concentrations de part et d'autre de la valeur du paramètre de toxicité ;
c) Le nombre de répétitions ;
d) Le nombre d'organismes par répétition ou concentration ;
e) La variation (dispersion) des points de données ; et, pour certaines méthodes d'analyse, la pente de la droite de régression.

L'expérimentateur peut tenter d'aménager les conditions favorables aux facteurs *a*) à *d*). Il est question, dans le présent paragraphe, du facteur *a*).

Si on pouvait prévoir un plan d'expérience idéal, celui-ci comprendrait plusieurs concentrations dans la gamme « médiane » des effets, avec une plage égale de concentrations au-dessus et au-dessous de cet intervalle. En essayant de choisir ces concentrations, l'expérimentateur est tiraillé entre deux partis :

(1) Resserrer les concentrations afin d'obtenir une bonne sélection d'effets « médians » ou partiels ;

(2) Disperser largement les concentrations, pour s'assurer qu'il y correspondra de petits et de grands effets.

Le problème de loin le plus fréquent découle de la première influence. L'expérimentateur est porté à choisir des concentrations relativement rapprochées, mais la valeur du paramètre de toxicité se trouve ailleurs que prévu. Comme nous y avons fait allusion, cela pourrait empêcher d'estimer le paramètre de toxicité dans un programme réglementaire ou de surveillance.

Ainsi, l'expérimentateur devrait résister à la tentation de choisir des concentrations trop serrées et il devrait prêter beaucoup plus attention qu'il ne l'aurait cru nécessaire à la seconde influence énumérée ci-dessus. Pour inspirer confiance, un ensemble de résultats devrait comprendre une faible concentration, qui provoque un effet semblable à celui que l'on observe chez le témoin, et une forte concentration, qui provoque un effet presque maximal. Ce plan d'expérience est préconisé par l'OCDE (OECD, 2004), selon qui « l'idée intuitive de concentrer les doses à proximité de la CE x n'est pas optimale. Les plans d'expérience englobant des niveaux de réponse suffisamment différents par rapport à ceux des témoins sont plus performants ».

D'autre part, en se laissant aller à la seconde influence, on risque de ne pas obtenir d'effets dans une gamme médiane de concentrations, c'est-à-dire des effets qui encadrent le paramètre de toxicité et qui déterminent les limites de confiance. Cela est un moindre mal. Au moins on saurait que le paramètre de toxicité se situe dans un intervalle donné, ce qui est mieux que de répondre qu'il est supérieur (ou inférieur) à une concentration *x*. Si on analyse les données par régression, une grande dispersion des données est un élément favorable. La régression est relativement impuissante, lorsque les données ne correspondent qu'à une partie de l'étendue des effets. Il importe de fixer les valeurs extrêmes de la distribution — une fois qu'elles sont établies, la partie médiane de la distribution suit, parce que la forme de la régression a habituellement été fixée par le choix d'un modèle.

On pourrait habituellement faciliter le choix des concentrations en commençant par des groupes supplémentaires d'organismes que l'on exposerait à des intervalles convenables de concentrations, de 6 à 8, comme il est recommandé dans certaines méthodes publiées par Environnement Canada, ou même plus. Si nécessaire, on pourrait utiliser moins d'organismes par concentration, tant qu'il y en aurait suffisamment pour satisfaire aux exigences minimales d'Environnement Canada. D'un point de vue statistique, il est préférable de disposer d'un plus grand nombre de concentrations (correspondant à des effets partiels ou « médians ») et d'y exposer (à chacune) moins d'organismes, plutôt que d'avoir plus d'organismes exposés à moins de concentrations. Comme nous l'avons mentionné ailleurs, on a prouvé que le fait d'employer 7 poissons au lieu de 10 aux concentrations appropriées ne diminuait pas gravement la précision de l'estimation du paramètre de toxicité (Douglas *et al.*, 1986).

Le fait de disposer de concentrations supplémentaires ne permettrait pas seulement une plus grande dispersion, mais, peut-être aussi, il permettrait d'avoir les petits intervalles souhaitables entre les concentrations.

2.2.2 Types particuliers d'essais

Les essais quantiques (de toxicité) correspondent à la distribution précédemment exposée. L'ensemble idéal de résultats serait constitué principalement d'effets partiels (aucun effet nul ni total), centrés sur la CE 50 et l'encadrant. La méthode recommandée d'analyse par la régression probit exige absolument deux effets partiels dans la série de résultats. Les effets près du taux de 50 % ont le plus d'influence dans l'estimation de la CE 50 et ils permettent de rétrécir l'intervalle de confiance.

Cependant, si le fait de disperser davantage les concentrations n'a permis d'obtenir qu'un seul effet partiel, on peut recourir à une deuxième méthode d'analyse (§ 4.5.6). Même sans effet partiel, on peut estimer la CE 50. Dans ce cas, les effets à 0 et à 100 % correspondrait à des concentrations successives de la série, probablement signe d'une CE 50 se situant dans un intervalle raisonnablement étroit (§ 4.5.7).

On pourrait **analyser** les résultats quantitatifs d'essais de toxicité sublétale **par régression.** Un aspect très

important du plan d'expérience serait de s'assurer que les observations couvriraient le domaine des petits à grands effets. Un modèle ajusté décrirait le mieux les trois phases suivantes : initialement, l'absence d'effet ou un effet faible, puis une région correspondant à un effet croissant, aboutissant à une région où l'effet est complet ou presque asymptotique, ne dénotant presque aucun changement. Il serait très souhaitable d'obtenir des données sur chacune de ces phases (§ 4.4 et 4.7). Pour ce qui concerne les essais quantitatifs de toxicité sublétale, certains exemples présentés dans le § 10.4 (« résultats difficiles ») montrent les incertitudes que soulève un plan d'expérience ne couvrant pas une étendue suffisamment grande de concentrations.

Pour la régression, il est généralement plus avantageux d'augmenter le nombre de concentrations plutôt que de répétitions. Bien que cela puisse ajouter au coût de mise en place de l'essai, cela pourrait ne pas accroître le nombre total d'organismes et cela pourrait même le réduire.

Les essais réglementaires comportent d'habitude moins de difficultés dans le choix des concentrations appropriées. Les essais en routine d'un effluent liquide exigeraient souvent une concentration à 100 %, qui fixerait visiblement la valeur maximale de la série de concentrations. On s'intéresserait probablement plus à l'intervalle de 1 à 100 % de concentration. En conséquence, une série commune et adéquate de concentrations dans des essais sur des effluents est de 100, 50, 25, 12,5 et 6,25 %. D'habitude, l'essai réglementaire préciserait un nombre minimal d'organismes.

Les essais menés à des fins de recherche ou d'enquête, par ex. pour déterminer la toxicité d'une substance nouvelle, pourraient exiger des efforts supplémentaires pour le choix des concentrations. Une *recherche préliminaire de la gamme de concentrations à utiliser* serait une tactique efficace, tant que les matières à l'étude sont stables pendant l'essai. On établirait ainsi les concentrations qu'il est rentable d'utiliser dans un essai définitif. Cet essai préliminaire pourrait employer un plan d'expérience tout à fait rudimentaire, ne comportant que quelques organismes ou concentrations ainsi qu'une exposition courte.

2.3 Logarithmes de la concentration

Dans la mise en place de l'essai, on choisit presque automatiquement une série de concentrations dont chacune multiplie la précédente par une constante. Ces concentrations forment ainsi une suite géométrique ou logarithmique. L'origine de cette règle est biologique et elle n'a rien à voir avec la statistique. C'est apparemment la façon dont les organismes « perçoivent » l'échelle des concentrations et, presque universellement, c'est ainsi que l'on réalise leur exposition. Faire autrement, c'est rendre l'essai moins efficace et moins puissant [1]. On pourrait utiliser une suite géométrique avec quelque multiplicateur que ce soit, par ex. une simple suite dont le multiplicateur serait 2, ce qui donnerait des concentrations telles que 2, 4, 8, 16, etc. Ou, si on divisait la fourchette voulue de concentrations par le nombre voulu d'enceintes expérimentales, on pourrait calculer un multiplicateur plus inhabituel, par ex. 1,6 (concentrations de 2, 3,2, 5,1, 8,2, etc.). Toute suite à multiplicateur constant aurait des intervalles logarithmiques égaux. Pour l'analyse, il est habituel d'exprimer les concentrations en logarithmes de base 10, mais on pourrait tout aussi bien utiliser les logarithmes népériens, tant que, à l'intérieur de l'essai, on utilise toujours les mêmes logarithmes.

Repères

- *Une suite géométrique (= logarithmique) de concentrations d'exposition est chose courante dans les essais de toxicité, pour de bonnes raisons. Une fois adoptée, elle reste la règle implicite, comme élément d'une bonne méthode scientifique. Après l'analyse statistique, on convertit habituellement les paramètres de toxicité et leurs limites de confiance en valeurs arithmétiques, pour les rendre plus intelligibles. Cependant, tout traitement mathématique ultérieur tel que le calcul de la moyenne et*

1. Par exemple, Robertson *et al.* (1984) ont constaté que des essais précis visant à déterminer la CL 50 exigeaient un espacement régulier des concentrations sur une échelle logarithmique, particulièrement les concentrations auxquelles correspondaient des effets de 25 à 75 %. Ils étudiaient particulièrement les besoins d'un essai efficace de toxicité, en utilisant des insectes dans les essais préliminaires.

l'analyse de variance devrait, par défaut, utiliser les logarithmes, à moins de prouver que ceux-ci ne conviennent pas.

- *Les programmes informatiques actuellement disponibles dans le commerce pour l'analyse des résultats sublétaux enfreignent habituellement le principe qui précède et, par défaut, ils emploient dans leurs calculs les valeurs arithmétiques de la concentration. L'expérimentateur doit comprendre le fonctionnement de ces programmes. Dans certains cas, la seule solution consiste à saisir les concentrations sous forme logarithmique.*
- *Les paramètres de toxicité calculés à l'aide de concentrations arithmétiques deviennent de plus en plus erronés à mesure que les ensembles de données deviennent plus variables.*
- *La durée d'exposition est également de nature logarithmique, et il faut utiliser le logarithme du temps dans les calculs.*

La plupart des gens comprennent intuitivement les motifs pour lesquels on met ainsi les essais en place, et on peut se représenter ces motifs en faisant appel au sens commun. L'emploi d'une suite arithmétique pourrait être acceptable à de faibles concentrations (par ex. 1, 2, 3, 4 et 5 mg/L), mais on ne voudrait probablement pas conserver le même intervalle arithmétique unitaire à de plus fortes concentrations (par ex. 11, 12, 13, 14, 15 mg/L). À des concentrations encore plus fortes, l'intervalle devient ridicule : pourquoi utiliser 101, 102, 103, 104 ... mg/L ? ou 1 001, 1 002, 1 003 ... mg/L ? Il serait probablement impossible de déceler une différence dans l'effet provoqué chez les organismes exposés à des concentrations de 101 et de 102 mg/L, sans parler de celles de 1 001 et de 1 002 mg/L. Le principe des intervalles plus étendus aux fortes concentrations se passe d'explication, par ex. 1, 2, 4, 8, ... ou 100, 200, 400... ou 1 000, 2 000, 4 000, etc.

Ainsi la modification des effets sur les organismes est reliée à l'augmentation proportionnelle de la concentration et non à son augmentation absolue. Bien qu'une augmentation de 10 unités, entre 10 à 20 mg/L, constitue un doublement de la concentration, la même variation arithmétique, de 100 à 110 mg/L, ne représente qu'une augmentation de 10 %, qui équivaut à une variation de 10 à 11 mg/L. Le doublement de la concentration toxique dans l'intervalle supérieur serait de 100 à 200 mg/L. Si ce dernier argument ne convainc pas, qu'il suffise de réfléchir aux unités de concentration : la suite 1, 2, 4, 8 mg/L est *identique* à la suite 1 000, 2 000, 4 000, 8 000 µg/L, même si la variation absolue semble énormément différente à première vue.

En conséquence, l'emploi du logarithme de la concentration reflète un phénomène biologique, il est adapté à l'exposition toxicologique et il *n'est pas* une transformation adoptée principalement pour faire plaisir aux statisticiens.

Il est parfois arrivé que les expérimentateurs qui utilisaient normalement une suite géométrique de concentrations, mais qui voulaient obtenir des renseignements plus détaillés sur une fourchette particulière étroite, comprise dans l'étendue de la suite, aient renoncé au principe géométrique et aient utilisé une suite arithmétique. Par exemple, dans une fourchette à laquelle ils s'intéressaient particulièrement, de 30 à 60, ils ont pu utiliser des concentrations de 30, 40, 50 et 60. Le même principe s'applique : il aurait fallu que cette fourchette soit couverte par une suite géométrique. En effet, les intervalles choisis à l'intérieur de la fourchette sont inégaux : celui de 30 à 40 représente une augmentation de 33 %, tandis que celui de 50 à 60 ne représente qu'une augmentation de 20 %. Habituellement, il aurait davantage convenu de diviser toute l'étendue des concentrations utilisées dans l'essai, y compris la fourchette à laquelle on s'intéressait le plus, en intervalles plus fins mais égaux.

On connaît peu d'exceptions à la transformation logarithmique de la dose. L'une d'entre elles serait le pH, qui est déjà logarithmique. Une autre serait la température, qui représente un cas particulier et, pour les interprétations appliquées au vivant, ne possède rien d'équivalent au zéro de l'échelle de concentrations d'un toxique.

2.3.1 Constance dans l'emploi des logarithmes

Bien que les expérimentateurs canadiens semblent adopter facilement une suite géométrique de concentrations d'exposition, ils sont souvent peu disposés à continuer d'employer les logarithmes pour

l'analyse statistique et, parfois, ils y sont franchement hostiles. La raison n'en est pas claire, mais cela pourrait s'expliquer par la complexité arithmétique accrue des logarithmes et/ou par leur méconnaissance (v. le § 2.3.5 sur la familiarisation). C'est une erreur fréquente que d'effectuer des analyses statistiques avec, pour la concentration, des valeurs arithmétiques. Si les résultats semble satisfaisants, les expérimentateurs ne voient pas le besoin d'utiliser les logarithmes et ils vont de l'avant avec les valeurs arithmétiques. Comme nous le décrirons plus loin, cette démarche est rétrograde, parce qu'elle consiste à abandonner l'échelle initiale sans exposer ses raisons. La bonne façon de faire consiste à entreprendre l'analyse avec les logarithmes des concentrations. Si on satisfait ainsi aux exigences de la méthode et si les résultats sont satisfaisants, on conserve l'échelle des concentrations.

L'expérimentateur ayant adopté l'échelle géométrique ou logarithmique pour l'essai et ayant ainsi rejeté le modèle arithmétique fait simplement preuve de rigueur scientifique en restant fidèle à son choix tout le long de l'étude et de l'analyse, à moins que l'échelle choisie ne se révèle erronée. L'adoption d'une échelle s'apparente à l'adoption d'une hypothèse — on y reste fidèle jusqu'à ce que l'on prouve que l'on a tort, auquel cas on cherche une meilleure hypothèse ou, dans ce cas, une meilleure échelle de concentration. Ce n'est pas d'abord une question de toxicologie ou de statistique, mais de science et de méthode scientifique. La suite géométrique de concentrations possède une « vérité » fondamentale. Dans son application (v. le 1er alinéa du § 2.3) et pour l'analyse statistique, on devrait conserver ce caractère fondamental si les effets doivent être interprétés sans altération. Même les statisticiens pourraient, de temps en temps, oublier cette raison fondamentale de conserver les logarithmes de concentration dans les analyses des résultats, et les biologiques ou les toxicologues devraient être prêts à défendre le concept. Dans des déclarations publiées, des statisticiens reconnaissent que le modèle d'analyse devrait suivre la « raison scientifique sous-jacente », ce que nous comprenons comme étant les motifs pour adopter une suite géométrique des concentrations d'exposition (par ex. Collett, 1991, p. 94)[2].

L'abandon de l'échelle géométrique à mi-chemin dans une étude est fréquent, malheureusement. Cela pourrait aider si l'expérimentateur se demandait pourquoi les expositions initiales ont été converties dans une suite logarithmique. *Les motifs de cette transformation, quels qu'ils soient,* restent valides durant les analyses statistiques, tant que l'on ne les a pas infirmés.

Si un test statistique formel a montré que le modèle n'est pas ajusté de façon satisfaisante aux données, il se peut alors que l'échelle logarithmique ne convienne pas [3]. L'ajustement pourrait être vérifié pour une transformation arithmétique des concentrations ou pour une autre sorte de transformation, afin de trouver un ajustement convenable (§ 2.9). Si l'échelle de concentrations de rechange s'est révélée supérieure, alors, pour le test, on aurait dû vraiment utiliser cette suite pour les concentrations d'exposition. Cependant, au risque de nous répéter, nous affirmons que ce sont les valeurs logarithmiques des concentrations qui représentent les valeurs par défaut et que les valeurs arithmétiques *ne doivent pas* être prises comme valeurs par défaut ni être utilisées d'abord pour la vérification du modèle.

La conservation des logarithmes de concentration signifie que toutes les manipulations mathématiques ultérieures des données devraient être à l'échelle logarithmique (§ 2.3.2 à 2.3.4).

2.3.2 Logarithmes et programmes informatiques

L'expérimentateur *doit* s'assurer, s'il utilise un programme informatique conçu pour les essais toxicologiques, que les calculs emploient les

2. Sur la question de l'emploi ou non des logarithmes dans l'analyse par la méthode des probits ou des logits, Collett s'exprime ainsi : « *Faute de motifs scientifiques sous-jacents* pour utiliser les valeurs transformées d'une variable explicative du modèle, le choix entre les modèles de rechange repose sur des bases uniquement statistiques, et on adoptera le modèle le mieux ajusté aux données disponibles. » (Les italiques sont de nous.) Bien que, à première vue, cette déclaration semble donner la primauté aux considérations statistiques, cela ne serait vrai que s'il n'y avait pas de motif scientifique (biologique) pour adopter un modèle donné.

3. L'expérimentateur ne doit pas oublier que l'absence d'ajustement du modèle aux données pourrait ne pas être attribuable à la transformation de la variable indépendante (la concentration), mais, plutôt, être attribuable à une ou à plusieurs des causes suivantes : *a*) la transformation de l'effet est nécessaire ; *b*) la distribution des tolérances (v. le glossaire) n'obéit pas à la loi normale ; *c*) la distribution des tolérances emploie différentes échelles (variances) à différentes concentrations ; *d*) le modèle choisi ne convient pas aux données.

logarithmes de la concentration.

Presque tous les programmes informatiques offerts sur le marché posent comme hypothèse que les concentrations d'exposition obéissent à une suite logarithmique, comme on peut le voir par leurs ensembles spécimens de données. La plupart des programmes conservent automatiquement les logarithmes pour la régression probit, mais ils ne les utilisent pas nécessairement pour d'autres types d'analyses des données. Les programmes diffèrent, et il pourrait être difficile de discerner quelle échelle de concentration ils utilisent. Inexplicablement, un vieux programme commercial avait apparemment la concentration arithmétique comme réglage par défaut ; un essai de TOXSTAT 3.5 a révélé qu'il en était ainsi, même pour la régression probit. Une option permettait de choisir le logarithme de la concentration, mais l'opérateur devait immédiatement saisir une autre commande, « RUN » (exécuter), sinon cette instruction était ignorée. Si, dans un programme commercial, toutes les autres tentatives échouent, on devrait saisir les concentrations sous forme de logarithmes. Tout tableur fournit les logarithmes, mais beaucoup de programmes commerciaux conçus pour l'analyse de la toxicité exigent que chaque élément de l'ensemble de données soit saisi (tâche fastidieuse) dans un segment particulier du programme.

Étonnamment, le programme ICPIN n'utilise pas de concentrations logarithmiques dans l'« interpolation linéaire » visant à estimer la CI p dans les essais quantitatifs de toxicité sublétale (§ 6.4 ; Norberg-King, 1993). Les procédures du programme ont initialement été mises en place par le personnel de l'USEPA et elles sont désormais intégrées dans les programmes commerciaux usuels.

Les programmes informatiques prenant en charge la nouvelle approche de la régression non linéaire sont d'usage général et ne sont pas conçus pour la toxicologie. Rien n'y est prévu pour qu'ils utilisent automatiquement le logarithme de la concentration, et des auteurs, en utilisant la régression non linéaire ont omis de transformer les valeurs arithmétiques (§ 6.5.7). Bien que la régression non linéaire puisse s'ajuster à presque toute forme de courbe, il y aura probablement une pénalité à payer (plus de paramètres à employer, avec perte de puissance pour l'ajustement) [voir le § 6.5.5]. Habituellement, la relation décrite par un modèle comportant des concentrations et/ou des durées sous forme logarithmique sera moins complexe. Le modèle peut s'ajuster à une relation plus simple sous forme de courbe ou de droite, avec moins de paramètres à estimer et perte moins grande de degrés de liberté, ce qui augmente la puissance de l'analyse. En outre, les courbes et les graphiques des effets sur une échelle arithmétique pourraient induire en erreur (§ 5.3).

2.3.3 Calculs ultérieurs avec des logarithmes

Une fois que l'on a calculé les paramètres de toxicité avec leurs limites de confiance, on convertit souvent ces résultats en valeurs arithmétiques pour les rendre plus intelligibles. Cependant, avant toute manipulation mathématique ultérieure des formes arithmétiques de ces valeurs ou des variables associées, il faut les reconvertir en logarithmes. (C'est le « *dose metameter* » [unité la plus pertinente pour exprimer la dose] de Finney, 1971.)

Une erreur fréquente consiste à calculer la moyenne des valeurs arithmétiques de deux ou de plusieurs CE 50, CI p ou autres paramètres de toxicité. Il faut se rappeler que, dans ces abréviations, *C* signifie *concentration*. On devrait se représenter le paramètre de toxicité estimé grâce à l'essai comme un logarithme parfois transformé temporairement en valeur arithmétique. La bonne façon de faire consiste à calculer la moyenne des valeurs logarithmiques des paramètres de toxicité, puis, si on le désire, à prendre l'antilogarithme du résultat (une moyenne géométrique) [4]. Quand on évalue les études de la toxicité, il reste nécessaire de dépister ce type d'erreur.

2.3.4 Cela importe-t-il ?

Des expérimentateurs protestent contre l'emploi de logarithmes, parce que les résultats sont semblables aux valeurs arithmétiques de la concentration. Bien que cela soit raisonnablement vrai pour quelques « bons » ensembles de données, il existe des différences appréciables dans les irrégularités souvent constatées en écotoxicologie.

4. Dans le document de travail pour l'élaboration de critères de qualité de l'eau (Stephan *et al.,* 1985), on trouvera un excellent exemple, fourni par l'USEPA, du moyennage poussé des paramètres de toxicité et d'autres manipulations des données à l'aide des méthodes logarithmiques appropriées.

Si les logarithmes donnent une estimation plus vraie des données irrégulières, cela conforte le principe selon lequel la bonne marche à suivre consiste à utiliser les logarithmes pour tous les ensembles de données. La question n'est pas : «Cela fait-il beaucoup de différence ? », mais plutôt : « Qu'est-ce qui est correct ? » Les expérimentateurs canadiens devraient utiliser la façon correcte.

Dans l'annexe D, on trouve deux exemples favorables à l'emploi des logarithmes. Le premier est simplement une comparaison des moyennes arithmétique et géométrique de certains ensembles de concentrations constituant des estimations de la toxicité. Dans le cas de données cohérentes (« bonnes »), l'écart entre ces moyennes était minime. Cependant les deux types de moyennes divergeaient de plus en plus à mesure que les ensembles de données devenaient plus irrégulières. Dans un cas extrême, la moyenne arithmétique était 5,4 fois plus élevée et elle n'était pas représentative de la plupart des valeurs de l'ensemble.

Dans le second exemple, on a calculé les CE 50 à l'aide des quatre ensembles de données A à D du tableau 2. La régression probit appliquée aux concentrations arithmétiques a donné des CE 50 qui étaient en moyenne 1,2 fois plus élevées que les valeurs justes. En général, les intervalles de confiance se sont élargies. On présente un autre exemple, avec une erreur d'importance semblable, du calcul erroné de la zone de confiance pour des toxiques de référence (§ 2.8).

Grâce à des mathématiques complexes, on pourrait réduire au minimum ce type d'erreur, mais *on ne pourra jamais supprimer la faille fondamentale de l'approche.*

Des erreurs semblables pourraient entacher les manipulations ultérieures qui n'auront pas utilisé le logarithme des paramètres de toxicité (par ex. moyennes, tendances au cours du temps, comparaisons de l'activité, analyse de variance, etc.). Elles pourraient aboutir à un classement erroné dans des catégories d'échec ou de réussite ou à la prise de mesures visant des écarts qui n'étaient pas réels. Une courbe de toxicité erronée pourrait faire conclure, à tort, à un seuil d'effet, comme le montre un exemple donné dans le § 5.2.

2.3.5 Familiarisation et techniques

Calculateurs électroniques et ordinateurs sont facilement disponibles depuis plusieurs décennies. À cause de cela les expérimentateurs d'aujourd'hui connaissent souvent mal les logarithmes et leur structure. Il peut être avantageux de consacrer un peu de temps à l'étude de leur nature. Une calculette possédant une touche « logarithme/antilogarithme » permettra de s'en faire une idée rapide. Il serait particulièrement intéressant d'étudier les manipulations arithmétiques par rapport aux manipulations des équivalents logarithmiques :

- Multiplication ou division correspondent à l'addition ou à la soustraction de logarithmes ;
- Les racines carrées et les autres racines fonctionnent par division des logarithmes.

Le glossaire donne d'autres explications, exemples à l'appui, sur la forme que prennent les logarithmes.

Si on est embarrassé d'utiliser une échelle logarithmique, on devrait songer que la concentration d'ions hydrogène dans l'eau est habituellement décrite par le pH, c'est-à-dire un logarithme, et que la plupart semblent s'en accommoder facilement.

L'emploi des logarithmes présente des difficultés et des inconvénients, mais on peut contourner ce problème. Ainsi, dans le cas des données relatives aux témoins, la saisie de la concentration peut faire problème parce qu'une concentration nulle ne possède pas de logarithme. Cela pourrait soulever des difficultés dans l'estimation de la CI p au moyen de la version actuelle du programme ICPIN, qui exige la saisie d'une concentration pour le témoin [5]. La solution consiste à saisir le logarithme d'une concentration quelconque, très faible par rapport à la concentration soumise à l'essai (disons 0,001 mg/L). En réalité, un programme tel qu'ICPIN n'utilise pas cette valeur, mais il reconnaît l'effet observé chez les témoins par sa position dans le tableau de données, de sorte que les procédures

5. Le programme ICPIN ne convertit pas les concentrations en logarithmes, de sorte que les utilisateurs canadiens doivent saisir les logarithmes de toutes les concentrations. Cela comprend la saisie du logarithme d'une très faible concentration, correspondant au témoin, dans la deuxième rangée du tableau de données.

d'analyse n'en souffrent pas.

Les concentrations inférieures à l'unité (1,0) ont un logarithme négatif, ce qui risque de porter à confusion. Pour les toxicologues, la solution la meilleure consiste à modifier les unités de concentration. Si, par ex., la concentration minimale à l'étude est de 0,1 mg/kg, et que, à la place, on utilise des microgrammes par kilogramme (µg/kg), les valeurs correspondantes partiront de 100 en montant, ce qui donnera des logarithmes positifs. Si les valeurs rebutantes étaient des concentrations exprimées en pourcentage, on pourrait en modifier l'échelle en millièmes ou en dix millièmes. Après les calculs, on pourrait convertir les résultats en valeurs arithmétiques, dans un souci d'intelligibilité, et modifier les unités selon son bon plaisir. Les programmes informatiques modernes manipulent sans difficulté les logarithmes négatifs, contrairement aux vieux programmes ou aux programmes « locaux ». Dans un souci de prudence, on devrait donc ne saisir que des logarithmes positifs dans les programmes informatiques.

2.3.6 Logarithme du temps

En écotoxicologie, le temps fait habituellement partie de la dose, de même que de l'effet ou de la réponse et, de ce fait, il faut également en tenir compte sous sa forme logarithmique. La nature géométrique/logarithmique du temps n'est pas si évidente en soi, mais on peut raisonner comme nous l'avons fait à l'égard des concentrations. Ce n'est pas la durée absolue qui détermine un changement d'effet, mais l'augmentation proportionnelle du temps. Dans un essai de toxicité, une augmentation de la durée d'exposition d'une heure à deux heures représenterait le doublement de la durée d'exposition, peut-être accompagnée d'une modification importante de l'effet. Si la durée passait de 96 à 97 heures, cela représenterait une augmentation insignifiante, probablement non décelable quant à la modification de l'effet. On devrait donc envisager l'emploi des logarithmes de temps dans le plan d'expérience et les utiliser dans toute analyse portant sur le temps.

Cela est quelque peu reconnu, car les expérimentateurs sont susceptibles de faire des inspections fréquentes au début d'un essai, puis, graduellement, ils les espacent. On reconnaît ainsi tacitement qu'une heure au début d'un essai d'une durée d'une semaine a plus d'importance qu'une heure à la fin. Les psychologues font observer que la perception humaine du temps écoulé est logarithmique dans une certaine mesure (Cohen, 1964). Un toxicologue spécialiste du milieu aquatique des premières années (Wilber, 1962) a décrit la situation comme suit :

> *« Temps biologique*
> *« Dans les longues études employant des concentrations sublétales de toxiques, il importe de reconnaître que le temps biologique est un phénomène logarithmique [Du Nouy, 1936]. D'autres ont rappelé ce fait [Gaddum, 1953]. Cela peut expliquer en partie pourquoi les courbes dose-réponse duquel le temps fait partie sont de nature logarithmique.*
>
> *« Il ne faut pas oublier le caractère logarithmique du temps biologique dans l'interprétation des résultats des expériences à long terme ayant employé des toxiques dans l'eau. Il est évident que la valeur et la signification biologiques d'un intervalle donné de temps ne seront pas les mêmes au début et à la fin d'une exposition chronique. Cela est si important que cela pourrait modifier les conclusions auxquelles on arrive*[6]*. »*

Dans les analyses toxicologiques, le temps intervient principalement dans les courbes de toxicité (section 5), mais, pour ce qui concerne les effets quantiques, il est avantageux d'estimer le temps correspondant à un effet de 50 % (TE 50, § 5.1).

2.3.7 Logarithme de l'effet ?

Dans les essais de toxicité, la variable indépendante peut parfois être de nature logarithmique et on devrait l'analyser comme telle. Cela pourrait survenir lorsque l'on mesure des effets quantitatifs. Par exemple, lorsque l'on calcule la CI p du gain de poids chez les organismes, on estime les variations proportionnellement au poids des organismes. On calcule la CI p comme un pourcentage désigné d'altération, de gêne ou de dysfonction, c'est-à-dire comme une réduction proportionnelle par rapport aux témoins. Autrement dit, la CI p étant affaire de proportion, les intervalles sont donc géométriques ou logarithmiques.

6. Du Nouy, que cite Wilber, a écrit une monographie sur le temps biologique, tandis que Gaddum est l'un des premiers géants de la pharmacotoxicologie. (Voir la rubrique « Références ».)

La plupart des arguments présentés dans le paragraphe sur l'emploi des logarithmes comme échelle par défaut de la variable indépendante (concentration) sembleraient également s'appliquer aux variables dépendantes quantitatives qui sont proportionnelles de nature (par ex. le poids). Cependant, on applique généralement le concept dans une seule situation, c'est-à-dire dans la transformation des données sur les effets, afin de répondre aux exigences de la normalité et de l'homogénéité de la variance (v. le § 2.9).

À part ce type de transformation que l'on effectue au besoin, le quotidien de l'écotoxicologie fait peu de cas de la notion de l'échelle proportionnelle des effets. Peut-être la question émergera-t-elle lorsque de nouvelles méthodes seront mises au point. L'un des statisticiens les plus rompus à la toxicologie (Slob, 2002) a adopté cette approche dans la modélisation des données quantitatives. Slob (2002) décrit ses hypothèses dans une régression non linéaire : « Par défaut, on pose que les mesures suivent une distribution log-normale. Par conséquent, le modèle de dose-réponse est ajusté à une échelle logarithmique, c'est-à-dire que le modèle comme les données sont transformés en logarithmes... Donc, les moyennes des groupes ne sont pas arithmétiques, mais géométriques... »

2.4 Randomisation

> *« La randomisation est quelque peu analogue à l'assurance : c'est une façon de se prémunir contre les éventuelles perturbations, lesquelles peuvent être graves. »* (Cochran et Cox, 1957)

Repères

- *Les tests statistiques reposent sur l'hypothèse selon laquelle toutes les variables auxiliaires d'un essai de toxicité sont aléatoires. La randomisation devrait donc toucher tous les aspects du plan d'expérience et des modes opératoires. Cela comprend la randomisation des récipients renfermant différentes concentrations, leur position dans le dispositif expérimental et la disposition des organismes dans les récipients.*
- *On peut supprimer une autre possibilité d'erreur systématique si l'observateur ne connaît pas l'identité des enceintes expérimentales.*
- *On présente, dans l'annexe E, des méthodes pratiques de randomisation.*

Dans les essais de toxicité comme dans les autres travaux expérimentaux, la randomisation est cruciale pour l'inférence statistique. Elle confère validité aux hypothèses expérimentales en abolissant toute corrélation potentielle entre les unités expérimentales. L'indépendance des observations autorise une estimation sans biais des effets et des traitements. Davis *et al.* (1998) ont conclu que l'affectation non aléatoire des organismes peut introduire un biais significatif dans les estimations de la concentration létale. Toute tentative raisonnable de randomisation a permis de supprimer le biais, mais on a obtenu le moins de variation dans les résultats grâce à une affectation complètement aléatoire.

La randomisation devrait s'imposer dans tous les aspects du plan d'expérience et du mode opératoire d'un essai de toxicité. Tout test statistique pose que toutes les variables contribuant aux données sont aléatoires, sauf la variable à laquelle on s'intéresse, qui, dans ce cas, serait l'agent ou les agents toxiques. Si l'une des variables auxiliaires n'est pas délibérément randomisée, il y aura automatiquement lieu de s'interroger sur la validité du traitement statistique. En omettant de randomiser un élément, on pose que ce dernier ne biaisera pas les résultats ni invalidera les tests statistiques, ce qui pourrait être vrai. Cependant, s'il devait causer un biais ou une invalidation, il n'y a habituellement aucun moyen de s'en s'en assurer après l'essai. La seule façon d'échapper à l'incertitude est de randomiser tous les facteurs possibles de contribution, à part les concentrations et la durée d'exposition, que l'on choisit pour faire partie de la « dose ».

Si une méthode d'essai d'Environnement Canada impose la randomisation, il faut signaler toute omission à cet égard, et cette omission pourrait invalider l'essai. Pour ce qui concerne les essais réglementaires qui pourraient servir dans des poursuites, une randomisation appropriée supprime un motif de critique de l'essai (et de l'expérimentateur) par une organisation de l'extérieur qui aurait voulu mettre les résultats en doute.

Devraient notamment être randomisés les éléments suivants :

- *Les récipients utilisés pour les concentrations,* dont la randomisation est rarement faite, mais qui devrait être pratiquée. Si un récipient a servi dans un essai antérieur, un report du toxique est possible, en dépit du nettoyage, et il pourrait influer sur le nouvel effet observé dans ce récipient. On pourrait même imaginer que de nouveaux récipients neufs aient un défaut ou un constituant occasionnel qui influerait de façon irrégulière sur les résultats d'un essai.

- *La disposition des récipients au hasard, dans la pièce, l'incubateur, etc.* est spécifiée dans la plupart des méthodes publiées par Environnement Canada. Il pourrait y avoir des différences dans les conditions accessoires, selon l'emplacement particulier des récipients. Parfois, on résiste à cette condition, parce qu'un dispositif expérimental irrégulier des concentrations et des répétitions est susceptible de causer des erreurs dans l'enregistrement des données (v. ci-dessous).

- *La disposition aléatoire des organismes dans les récipients peut avoir de l'importance.* Souvent, on omet de se plier à cette condition parce qu'elle peut être ennuyeuse et que, parfois, il est difficile de conserver la trace du nombre d'organismes ayant été déposés dans un récipient donné. La randomisation formelle est possible et même un système comme celui de la distribution de cartes peut être satisfaisant.

- *Les essais en aveugle, dans lesquels l'observateur ne connaît pas les traitements, signifient que les récipients doivent être identifiés par un code plutôt que par l'indication de leur concentration.* L'essai en aveugle dénote le souci poussé d'éviter le biais dû à l'observateur et il contribuerait à rendre inattaquables ses résultats.

La pire situation serait un *biais* (erreur systématique) dû à l'omission de la randomisation. Par exemple, si on a attribué les concentrations aux organismes de la colonie dans l'ordre de leur capture, les organismes capturés le plus facilement pourraient correspondre aux concentrations minimales. Il se pourrait que ces organismes soient plus faibles et plus sensibles aux toxiques, ce qui exagérerait l'effet du toxique à faible concentration. De même, si les enceintes expérimentales étaient alignées dans l'ordre des concentrations, les résultats pourraient être biaisés par un gradient de température, d'éclairage ou le gradient d'une perturbation qui existait dans le laboratoire. Par exemple, la proximité d'un appareil de chauffage dans un incubateur pourrait influer sur les températures de l'essai et, de là, sur la toxicité. Les essais avec des algues peuvent être particulièrement variables, parce que la croissance des algues chute quand l'éclairage diminue, ce qui pourrait se produire sur les bordures et dans les angles du dispositif expérimental. Même en dépit d'une excellente randomisation, des facteurs extérieurs méconnus pourraient influer sur la toxicité dans certains récipients, *mais cela augmenterait simplement la variation générale des résultats de l'essai, sans constituer une erreur systématique.*

Une randomisation compliquée pourrait contribuer à un risque défini d'erreur due à l'expérimentateur dans l'affectation des expositions ou l'enregistrement des données. Cela pourrait certainement augmenter le travail et sa durée. Même les statisticiens de l'OCDE (OECD, 2004) reconnaissent que, dans certaines circonstances, il peut être difficile ou coûteux de randomiser toutes les étapes d'une expérience. Si une partie de la randomisation doit être omise, ils recommandent l'examen séparé de l'effet potentiel de cette omission sur les résultats de l'essai. C'est pourquoi certains essais effectués au Canada ne sont probablement pas suffisamment randomisés, et les expérimentateurs devraient se rendre compte que leurs résultats risquent d'être biaisés. Si, pour de bons motifs, on altère une partie de la randomisation, on devrait le faire de façon que *seulement la variation totale de l'essai soit susceptible d'être modifiée,* en s'efforçant de réduire au minimum le risque de biais relié aux concentrations. La seule façon de s'assurer d'éviter ce biais est de randomiser totalement chaque étape de l'essai.

Dans l'annexe E, nous donnons des conseils utiles sur la distribution des organismes dans les récipients et le positionnement des récipients. La plupart des manuels de statistique offrent des conseils et des méthodes (par ex. Fleiss, 1981).

2.5 Répétitions et nombre d'organismes

Repères

- *Grâce à la répétition, on peut estimer la variation correspondant à chaque concentration, ce qui, à son tour, peut servir à se prononcer sur les différences significatives entre les concentrations.*
- *Dans un essai donné, la répétition doit être une enceinte expérimentale indépendante renfermant un ou plusieurs organismes, qui n'a pas de connexion avec une autre enceinte par le milieu d'essai.*
- *Un traitement comprend toutes les répétitions à une concentration donnée et tous les organismes dans chacune de ces répétitions.*
- *Il importe de bien comprendre et de bien utiliser la terminologie, sinon les tests statistiques risqueraient d'être utilisés d'une façon invalide.*
- *Le nombre d'organismes par concentration ou répétition est un facteur important dans le plan d'expérience. Dans le laboratoire, des limites pratiques peuvent empêcher l'emploi d'un nombre suffisant d'organismes pour atteindre les idéaux statistiques. Les répétitions pourraient constituer une façon avantageuse d'offrir des conditions convenables aux organismes en expérience ou de pouvoir compter une enceinte expérimentale de rechange en cas d'accident.*
- *Pour l'analyse des résultats d'un essai quantique par la régression probit, on réunirait toutes les répétitions. Cependant, les répétitions sont utiles lorsque l'on emploie des outils statistiques plus perfectionnés.*
- *Si la régression a servi à une estimation ponctuelle avec des données quantitatives, les répétitions permettent de tester la qualité de l'ajustement et l'écart du modèle par rapport aux données. On peut faire des estimations ponctuelles par lissage et interpolation, sans répétition, mais le programme ICPIN, utilisé communément, exige deux répétitions et, de préférence, cinq pour évaluer la signification. Les répétitions font partie intégrante des tests d'hypothèse.*
- *Lorsque l'on prélève des échantillons en vue d'essais, les échantillons réitérés de terrain (vraies répétitions) sont des échantillons séparés de sédiment, d'eau, etc. prélevés au même moment et dans le même emplacement général. Dans un essai de toxicité, elles constituent d'excellentes répétitions pour tenir compte de la variation du substrat que l'on évalue. Les sous-échantillons d'un échantillon (« répétitions de laboratoire ») permettent d'estimer la variabilité de la technique de laboratoire et l'homogénéité de l'échantillon, mais ils ne livrent aucun renseignement permettant de distinguer les emplacements sur le terrain.*

2.5.1 Terminologie

L'emploi de la bonne terminologie des essais de toxicité peut avoir de l'importance. Un usage fautif pourrait entraîner une mauvaise application d'un test statistique, risquant d'aboutir à des conclusions invalides.

Dans un essai de toxicité, une *répétition* est une enceinte expérimentale simple renfermant un ou plusieurs organismes et c'est l'une des enceintes (au moins deux) exposées au même *traitement,* c'est-à-dire exposées à la même concentration de matière à l'étude (ou exposées aux conditions témoins) [7]. Ainsi les répétitions répètent *l'unité expérimentale,* le plus petit élément indépendant à qui, dans un essai de toxicité, on applique un traitement. Cette terminologie est expliquée dans le texte qui suit ainsi que dans le glossaire.

Il pourrait n'y avoir qu'un seul organisme dans l'enceinte expérimentale, et cela constituerait quand même une répétition et une unité expérimentale. Un exemple est donné par l'essai de survie et de reproduction de *Ceriodaphnia* d'Environnement Canada (EC, 1992a). Chacun des 10 organismes géniteurs exposés à un traitement est une répétition et, également, une unité expérimentale parce qu'il se trouve dans une enceinte expérimentale séparée. Dans l'essai on dénombre le nombre de jeunes issus de chaque organisme.

7. Dans le présent paragraphe, les exemples de *traitement* concernent tous des concentrations, mais cela n'est pas obligatoire. Un échantillon de sédiment prélevé sur le terrain pourrait également constituer un traitement quand il a été soumis à un essai.

Cependant, comme les individus manifestent une sensibilité différente, un seul organisme par répétition signifie que les répétitions sont aussi variables que les organismes (d'où le nombre élevé de 10 répétitions dans l'essai avec *Ceriodaphnia*). On utilise normalement plusieurs organismes par enceinte pour améliorer la précision. Les organismes dans une enceinte sont des *unités d'échantillonnage* livrant des données qui contribuent au résultat relatif à la répétition.

La répétition doit être indépendante. Les enceintes séparées, qui sont des répétitions, ne doivent avoir aucune connexion entre elles par l'eau, le sédiment ou le sol d'essai. Ainsi, si plusieurs enceintes perméables utilisées dans un essai en milieu aquatique étaient exposées par suspension dans un bassin de solution d'essai, les enceintes ne constitueraient pas des répétitions. De même, la matière à l'étude étant entrée en contact avec une enceinte-répétition ne doit pas être transvasée pour entrer en contact avec une autre enceinte. Il ne peut pas y avoir de transfert d'organismes entre des enceintes, une fois que l'essai a débuté. Le non-respect de ces exigences invaliderait l'analyse statistique fondée sur les répétitions.

En écotoxicologie, une partie de la terminologie est flottante. L'expression « traitement répété » (*replicate treatment*) figure dans certaines méthodes publiées par Environnement Canada et elle possède la même signification que celle que nous attribuons à *répétition* (*replicate*). *Replicate treatment* est une expression déroutante, parce que chacun des mots qui la composent appartient à deux niveaux hiérarchiques différents (v. le texte qui suit), et nous recommandons plutôt d'employer le mot *répétition* (*replicate*). Parfois, des statisticiens utilisent le terme *répétition* pour désigner une enceinte expérimentale (Snedecor et Cochran, 1980) et, de la sorte, ils pourraient parler de plusieurs « répétitions » pour une concentration donnée, signifiant ainsi que plusieurs enceintes correspondaient au même traitement. Le mot « répétition » semblerait mieux utilisé comme mot d'action signifiant l'action de créer des répétitions.

L'expérimentateur devrait être à l'affût de toute erreur dans les instructions de logiciels où, parfois, on a désigné les organismes individuels dans une enceinte expérimentale sous l'appellation de « répétitions ». C'est une erreur, que l'on pourrait assimiler à une forme de *pseudo-répétition.* Ces organismes sont des *unités d'échantillonnage* ou des *sous-échantillons* contribuant à une répétition. En langage courant, l'information tirée d'un organisme pourrait être appelée « mesure » ou « observation » : par ex. « les 10 mesures effectuées dans la première répétition ont été... ». Des commentaires supplémentaires sur l'erreur de *pseudo-répétition* suivent sous la rubrique « Test d'hypothèse » (v. aussi le § 7.2.1).

Grâce à l'exemple d'un essai ordinaire de toxicité sublétale en milieu aquatique, on peut préconiser la bonne terminologie suivante :

4 concentrations d'essai et 1 témoin	= 5 *traitements*
2 enceintes expérimentales isolées pour chaque concentration	= 2 *répétitions* par traitement
6 poissons dans chaque enceinte	= 6 *unités d'échantillonnage* par répétition
En tout, 5 traitements comptant 2 répétitions	= 10 *unités expérimentales*

En conséquence, une expérience peut donner lieu à trois niveaux de variation dans les mesures :

- Sur chaque organisme d'un récipient (*les unités d'échantillonnage*) ;
- Entre chaque récipient se trouvant à la même concentration (*répétitions*) ;
- Sur les concentrations (*traitements*).

Manifestement, l'expérimentateur doit comprendre les différences, particulièrement quand il effectue une analyse de variance.

2.5.2 Répétition dans les diverses sortes d'essais

La répétition des enceintes expérimentales peut être une façon puissante d'améliorer la qualité de l'information tirée de certains essais de toxicité. Elle permet d'évaluer la variation ou le « bruit » correspondant à

chaque concentration et d'effectuer un test statistique du manque d'ajustement. On recommande chaudement la lecture de la communication de Hurlbert (1984) sur la répétition.

Répétitions dans essais quantiques. — Des répétitions à chaque concentration ne sont habituellement pas nécessaires, parce qu'on réunit tous les résultats correspondant à chaque concentration avant l'estimation de la CL 50 ou de la CE 50 par les méthodes classiques comme la régression probit, utilisée communément aujourd'hui. Les répétitions sont parfois commodes ou utiles, toutefois, pour manipuler les conditions convenant aux organismes en expérience et les leur assurer. Par exemple, le fait de répartir le nombre total d'organismes exposés à une concentration donnée entre plusieurs répétitions serait une façon de fournir le volume nécessaire de matière à l'étude dans un récipient de taille commode.

En outre, l'essai bénéficierait d'une « assurance » réelle en cas d'accident à une enceinte, de perte ou de maladie. Si une répétition subissait un tel malheur, les autres pourraient habituellement servir à l'analyse des résultats. Par exemple, Environnement Canada exige trois répétitions dans l'essai de toxicité sublétale ou létale avec les premiers stades du développement de la truite arc-en-ciel (EC, 1998a). L'essai ne semblerait pas exiger de répétitions, parce qu'il cherche à estimer la CE 50 et la CE 25, concentrations auxquelles correspondent la non-viabilité et le retard du développement. La raison en est qu'il existe un risque appréciable de dommages ou de maladie, par suite de la manipulation des œufs fragiles et des jeunes stades du développement de la truite et que les répétitions augmentent la probabilité d'obtenir des données convenables à chaque concentration [8].

Des répétitions sont utiles si on applique des programmes statistiques plus sophistiqués aux essais quantiques. L'utilisation de tels programmes pourrait se généraliser.

Nombres d'organismes dans les essais quantiques à répétition unique. — En augmentant le nombre d'organismes en expérience on peut améliorer la précision de l'essai, ce qui permet l'estimation d'un intervalle plus étroit de confiance pour le paramètre de toxicité. Dans les essais quantiques, on pourrait diminuer de moitié le quotient entre la limite de confiance et la CE 50, grâce à l'emploi de 30 organismes par enceinte expérimentale au lieu de 10 (Hodson *et al.*, 1977). Une amélioration semblable a été quantifiée par Jensen (1972), qui a constaté une diminution importante de la variance de la CL 50 alors que le nombre d'organismes est passé de 1 à 10 par traitement. En outre, il y a eu diminution de 29 % de l'erreur type lorsque le nombre d'organismes est passé de 10 à 20, de 13 % lorsque leur nombre est passé de 20 à 30, et de seulement 8 %, lorsqu'il est passé de 30 à 40. Les améliorations étaient modestes lorsqu'il y avait plus de 30 organismes par traitement dans ces essais de températures létales. Bien sûr, les résultats exacts de comparaisons comme celles-là dépendront de l'espacement des concentrations autour de la CL 50.

Les statisticiens recommandent vivement d'augmenter le nombre d'organismes pour améliorer la précision, mais d'autres facteurs influent aussi sur le choix du nombre d'organismes, par ex. le souci d'économie, la taille des récipients, le volume disponible d'échantillon et les lois sur les droits des animaux. Dans les essais avec le poisson, la tendance est d'en utiliser moins par enceinte expérimentale, en partie pour sacrifier moins d'organismes. Douglas *et al.* (1986) ont mentionné une petite perte de précision par suite de la réduction de 44 % du nombre d'organismes, c'est-à-dire par l'emploi de sept animaux dans chacune des quatre concentrations utilisées, au lieu de dix dans chacune des cinq prévues. Cependant, la réduction du nombre de concentrations expose l'expérimentateur au risque de manquer l'importante plage d'effet (§ 2.2) et, à n'en pas douter, la précision diminue quelque peu si le nombre d'organismes diminue à moins de 10 par traitement, comme nous l'avons mentionné dans l'alinéa précédent.

Estimations ponctuelles par régression. — Dans la tendance actuelle à employer la régression pour estimer les paramètres d'une toxicité se manifestant par un effet sublétal quantitatif (§ 6.5), il peut être avantageux d'utiliser des concentrations supplémentaires, plus

8. Des répétitions sont exigées pour d'autres essais d'estimation de paramètres de toxicité tels que l'essai avec des têtes-de-boule (EC, 1992b). Parce que les essais permettent de mesurer un double effet, avec effet létal et des effets sublétaux, les répétitions sont exigées pour le dernier.

rapprochées les unes des autres (Moore, 1996 ; § 6.2.3). En conséquence, on est encouragé à utiliser les ressources pour plus de concentrations plutôt que pour plus de répétitions. En effet, la régression classique exige, à proprement parler, une seule mesure à chaque concentration. Dans sa forme la plus simple, la *régression* décrit la relation linéaire entre une observation, telle que la taille, et une variable continue indépendante, telle que le logarithme de la concentration. Après avoir défini mathématiquement la relation, on s'en sert pour calculer le paramètre de toxicité. Les limites de confiance de ce paramètre peuvent être obtenues avec ou sans répétitions.

Néanmoins, des raisons majeures militent en faveur d'un grand nombre de répétitions. Environnement Canada a recommandé de 3 à 10 répétitions ou plus, dans des méthodes récemment publiées, qui exigent l'application de techniques de régression (EC, 2004a, b et 2007 et annexe O). La raison principale en est que des répétitions sont indispensables à l'évaluation de l'ajustement d'une régression[9]. Sans répétitions, il n'y a pas moyen de distinguer l'erreur due à la dispersion des observations à la même concentration (appelons cela l'*erreur pure*) d'une dispersion réelle des données due à la configuration du modèle (appelons cela l'*erreur due au manque d'ajustement*)[10].

Lissage et interpolation. — Si on doit se servir de la méthode du programme ICPIN (§ 6.4) pour estimer la CI p, il faut au moins deux répétitions pour calculer les limites de confiance. Chaque répétition contribue à une mesure, par exemple, le poids moyen des organismes dans cette répétition. Cinq mesures (répétitions) ou plus par concentration réduiraient la largeur de l'intervalle de confiance.

Test d'hypothèse. — Les répétitions sont essentielles à l'analyse des résultats par le *test d'hypothèse*, jadis une démarche privilégiée (section 7). Plus les répétitions sont nombreuses, plus elles favorisent l'analyse de variance, en permettant de distinguer avec plus de certitude la CSEO de la CEMO. Si l'expérimentateur a l'intention d'effectuer un test d'hypothèse tout en faisant une estimation ponctuelle, il pourrait ajouter plus de répétitions dans le plan d'expérience. Environnement Canada exige au moins quatre répétitions si l'on doit estimer la CSEO et la CEMO dans l'essai de toxicité sublétale employant des stades juvéniles de la truite arc-en-ciel (EC, 1998a). Ces quatre répétitions pourraient être essentielles si la statistique paramétrique était invalide et que l'on devait faire appel à des méthodes non paramétriques.

Le test d'hypothèse présente le risque particulier de pseudo-répétition (v. le § 2.5.1). Il n'est pas difficile d'imaginer les erreurs grossières qui pourraient entacher les conclusions si les organismes se trouvant dans une enceinte étaient, par erreur, inscrits comme répétitions dans une analyse de variance. Si, par exemple, 10 vers se trouvaient dans chaque enceinte, le test statistique considérerait à tort que l'expérience est puissante, en effet. D'après l'analyse, les différences aléatoires pourraient paraître significatives (« réelles »).

2.5.3 Relations avec l'échantillonnage sur le terrain

Quand des échantillons prélevés sur le terrain sont examinés au laboratoire, il y a des relations entre les modes opératoires et l'interprétation de résultats, compte tenu du terrain. Cela serait particulièrement approprié lorsque des échantillons de sédiment ou de sol (« substrat ») étaient apportés au laboratoire, mais, parfois, cela s'applique aux échantillons d'eau[11].

9. Dans certaines méthodes publiées antérieurement par Environnement Canada, il n'était pas impératif de vérifier la qualité de l'ajustement. La décision en était laissée à l'expérimentateur, s'il voulait montrer que le modèle de régression était convenablement ajusté.

10. Un avantage important de la répétition est de permettre de distinguer entre deux catégories de variations dans un essai donné. L'*erreur pure* serait la dispersion apparemment aléatoire causée par les sensibilités différentes de chaque organisme à la même concentration. L'autre catégorie serait l'*erreur due au manque d'ajustement,* c'est-à-dire des variations homogènes par rapport au modèle de régression choisi. Les répétitions sont nécessaires pour distinguer ces deux catégories de variations.

Un exemple d'*erreur due au manque d'ajustement* serait d'adopter une droite comme modèle supposé de la relation concentration-effet, alors que les données représentent une courbe convexe s'écartant de plus en plus de la linéarité aux fortes concentrations.

11. Le présent document n'offre pas de conseils sur le travail de terrain, mais des observations supplémentaires sur le prélèvement d'échantillons sont de mise pour l'organisation et l'interprétation des essais de toxicité. Parfois, décider de ce qu'est une répétition dans un échantillonnage sur le terrain, par exemple dans l'évaluation des sédiments d'une baie, est entouré d'une incertitude considérable. Le principe général est le suivant : des échantillons réitérés devraient couvrir convenablement la surface considérée comme uniforme, que l'expérimentateur souhaite

En particulier, il existe une différence très importante entre les répétitions d'un essai fondées sur des échantillons séparés de la matière à l'étude et les répétitions fondées sur la subdivision d'un échantillon (sous-échantillons). Les échantillons qui étaient des *échantillons réitérés* constitueraient des échantillons séparés de sol, de sédiment, etc., prélevés sur le terrain par des méthodes identiques et dans la même station d'échantillonnage. Leur objet serait de permettre l'évaluation de la variation de la qualité (ou des qualités) du substrat échantillonné à cette station. Ce type d'échantillon est parfois aussi appelé *échantillon réitéré de terrain.* Les échantillons réitérés doivent être gardés dans des récipients séparés et, comme cela est souvent recommandé pour les essais de sols ou de sédiments, chacun peut être utilisé pour constituer une répétition de chaque traitement dans un essai de toxicité. Le mode opératoire intégrerait dans l'essai de toxicité les variations suivantes combinées : *a*) la variation du sédiment ou du sol dans une station donnée (et la variation des modes d'échantillonnage) ; *b*) toute variation due aux conditions ou aux modes opératoires du laboratoire.

On pourrait créer des sous-échantillons au laboratoire en subdivisant un échantillon de substrat. Ces sous-échantillons sont également appelés *répétitions de laboratoire,* mais le terme « sous-échantillon » décrit bien leur nature. Si ces sous-échantillons étaient utilisés comme répétitions dans un essai de toxicité, les résultats permettraient d'estimer l'homogénéité de chaque échantillon et la variation due au mode opératoire, ce qui pourrait être une qualité du plan d'expérience. Cependant, les sous-échantillons *ne diraient rien sur la variation du substrat sur le terrain* (par ex. un sédiment lacustre ; v. le § 3.1.3). Selon le but de l'étude, il pourrait être plus rentable de faire porter l'effort consacré à la préparation et à l'analyse de sous-échantillons à l'obtention d'échantillons réitérés. En conséquence, nous ne recommandons pas habituellement l'analyse de sous-échantillons de laboratoire, à moins que ceux-ci ne facilitent la manipulation des organismes (enceintes moins peuplées), qu'ils n'aident à l'organisation des essais (par ex., récipients plus petits), qu'ils ne répondent à un besoin d'évaluer l'homogénéité de l'échantillon et la variation dans la technique de laboratoire ou qu'ils ne soient exigés par une méthode d'essai particulière.

Environnement Canada (1994) est un excellent guide sur l'obtention d'échantillons répétés de sédiments.

2.6 Pondération

Repères

- *La pondération de certaines observations leur donne plus d'influence (de poids) sur les résultats des calculs ultérieurs.*
- *On accorde plus de poids à une valeur, pour l'une des raisons suivantes :* a*) elle est proche du paramètre de toxicité auquel on s'intéresse ;*

caractériser. Si toute la baie doit être caractérisée en tant qu'unité, alors les échantillons prélevés en un certain nombre de points autour de la baie seraient des *échantillons réitérés* (de terrain). Dans ces circonstances, si on prélève un certain nombre d'échantillons en un seul point, ces échantillons ne seraient pas réellement des échantillons réitérés représentant la variation de toute la baie, mais, plutôt, des sous-échantillons d'un emplacement particulier dans la baie.

D'autre part, si l'expérimentateur voulait évaluer les effets de pollution dans différentes parties de la baie, la stratégie d'échantillonnage serait plutôt différente, tout comme la perspective sur les répétitions. Il pourrait y avoir un ensemble d'échantillons que l'on prélèverait dans une station d'échantillonnage dans le fond de la baie, près d'une source ponctuelle de pollution. Un autre ensemble pourrait être prélevé dans une station située dans la partie nord de la baie, ouverte sur le large, afin d'évaluer l'effet de la dilution que subit l'effluent dans son transport vers le large par un courant faisant le tour de la baie. Une troisième ensemble d'échantillons pourrait être prélevé dans une station du sud de la baie ouvert sur le large où on s'attendrait à ce que l'eau nouvelle pénétrant dans la baie ne soit pas polluée. Si plusieurs échantillons de sédiment étaient prélevés à chaque endroit, les échantillons d'une station seraient des répétitions. Il s'agirait de déterminer si les trois stations diffèrent quant à leur pollution, de façon significative par rapport aux variations mesurées grâce à la répétition des échantillons de chaque station.

Manifestement, pour être valides, les conclusions de l'étude de la toxicité exigeraient que l'échantillonnage sur le terrain se fonde sur une bonne compréhension des facteurs physiques agissant dans l'habitat auquel on s'intéresse. Par exemple, dans la baie susmentionnée, il pourrait y avoir différents mouvements d'eau en profondeur et près de la surface. Dans tout plan d'échantillonnage des sédiments, il faudrait traiter les deux profondeurs comme des zones différentes, en sus des zones réparties horizontalement par rapport à la baie.

Ces distinctions concernant les répétitions sont en rapport avec les programmes canadiens de Suivi des effets sur l'environnement, dans lesquels les études sur le terrain sont coordonnés avec les essais de toxicité au laboratoire.

b*) elle représente de nombreux organismes ou de nombreuses mesures ;* c*) elle représente des mesures dont la variation est petite.*

D'après le glossaire, la *pondération* d'un élément d'une série, signifie que l'on manipule arithmétiquement cet élément pour en modifier l'influence sur le calcul appliqué à la série. Les motifs communément invoqués pour justifier la pondération seraient le nombre inégal de mesures dans les groupes d'une série ou les variances inégales affectant les éléments d'une série. Les emplois de la pondération sont l'objet d'un plus long développement dans les alinéas qui suivent.

Un exemple de pondération est donné dans le § 4.2.3, sur l'utilisation d'un graphique tracé à la main pour estimer la CE 50. On y lit le conseil suivant : « en ajustant la droite des probits à vue d'œil,... on devrait pondérer mentalement les points. On devrait affecter du plus grand coefficient de pondération les points les plus rapprochés de l'effet de 50 %... ». D'un point de vue pratique, on pondère les valeurs centrales parce que cette plage de l'ensemble de données est la plus rapprochée du paramètre de toxicité auquel on s'intéresse et, fort probablement, parce que l'on veut l'estimer exactement. Ce type de pondération informelle est subjectif, c'est le moins qu'on puisse dire, mais cela est mieux que d'ignorer la valeur relative des points portés sur le graphique.

On pourrait introduire la notion de pondération formelle par un exemple simpliste d'ajustement mathématique d'une courbe. Si les valeurs que l'on estime être les plus importantes étaient saisies deux fois dans l'ensemble de données, elles auraient plus d'influence sur l'ajustement, c'est-à-dire qu'elles pèseraient plus lourd sur ce dernier. (Inutile de préciser que cela n'est aucunement une méthode autorisée et que nous la mentionnons uniquement pour exprimer l'idée de ce qu'est la pondération.) La pondération formelle est souvent une opération tout à fait sophistiquée, comme dans la régression probit (§ 4.5), où elle se fonde sur le probit prévu et elle possède une grandeur sans cesse variable.

La pondération sert communément à équilibrer le nombre de mesures contribuant à une valeur donnée dans une série. Si chaque valeur de la série était la moyenne de mesures effectuées sur un échantillon d'organismes, on pourrait pondérer une moyenne particulière parce qu'elle se fonderait sur un gros échantillon d'organismes. L'opération précéderait l'analyse. Le coefficient de correction pourrait être aussi simple que le nombre d'organismes.

On pourrait également pondérer la moyenne d'un groupe d'observations parce qu'elle provient d'observations ayant présenté une petite variation, ce qui fait que la moyenne semble une estimation particulièrement utile d'une série. Si le groupe d'observations lui-même était utilisé dans l'analyse, on pourrait lui appliquer directement la pondération. Ce type de pondération est indispensable lorsque l'on ajuste un modèle aux données dans lesquelles certains groupes sont plus variables que d'autres. Le modèle exigera presque certainement des variances égales (équivariances). On peut pondérer les observations selon une méthode mathématique intelligente pour que l'hypothèse de l'équivariance soit restaurée ; habituellement le programme informatique du modèle s'occupe de cette étape.

Nous faisons spécifiquement allusion à la notion de pondération dans les paragraphes suivants : 4.2.2, 4.2.3, 4.5.1 à 4.5.3 (divers aspects de la régression probit) ; 4.5.6 (estimations obtenues par la méthode de Spearman-Kärber) ; 4.7 (modèles non linéaires pour données quantiques et lissage pour les méthodes du noyau) ; 6.4 (détermination de la CI p par lissage et interpolation) ; 6.5.4 (inverse de la variance, pour la régression non linéaire) ; 8.2.3 (nombres inégaux de répétitions dans les essais de mesure d'un double effet).

2.7 Témoins

Dans un essai de toxicité, les témoins représentent un traitement soumis à tous les facteurs physicochimiques et biologiques qui pourraient influer sur les résultats de l'essai, sauf la condition précise à l'étude. Aucune des matières dont on étudie la toxicité n'est ajoutée au témoin. Le témoin sert de point de comparaison pour les effets expérimentaux résultant des conditions telles que la qualité de l'eau de dilution ou la santé et la manipulation des organismes. *Témoin* est synonyme de *témoin négatif.*

Une méthode pourrait exiger la répétition de chaque concentration d'essai. Dans ce cas, il faut répéter de

même chaque type de témoin. Certaines méthodes spécifient différents nombres de répétitions pour le toxique et le témoin (tableau O.1 de l'annexe O).

Repères

- *Le témoin doit être identique en tout aux concentrations d'essai, sauf qu'il ne renferme aucune des matières dont on étudie la toxicité (c'est-à-dire traitement à concentration nulle). Le témoin constitue le point de comparaison pour les effets que l'on observera.*
- *Si on se sert d'un solvant pour solubiliser la substance chimique à l'étude, il faut utiliser un témoin du solvant, qui renferme la concentration maximale de ce solvant employée dans l'essai. Ce témoin ne doit pas causer un effet plus grand que celui du témoin ordinaire.*
- *Si, dans un essai en milieu marin, la salinité n'est pas ajustée, il faut des témoins de la salinité correspondant à la teneur en sel des différents traitements. Si la salinité des eaux d'essai est ajustée à une valeur favorable (30 ‰), ce doit également être celle du témoin. Des témoins supplémentaires de la salinité sont nécessaires si les eaux d'essai sont ajustées par des méthodes (sel sec ou saumure) qui diffèrent de la méthode employée pour le témoin.*
- *Les essais effectués sur des sédiments et des sols utilisent des témoins qui obéissent aux mêmes principes que ceux qui s'appliquent aux autres essais. On compare les résultats des essais, normalement, à ceux qu'ont donnés un sédiment ou un sol de référence, prélevés sur le terrain et réputés non pollués. On emploie également un sédiment ou un sol témoin pour juger de la qualité globale de l'essai ; il s'agit d'un témoin artificiel ou d'un témoin prélevé dans un emplacement différent et non pollué.*

2.7.1 Témoins ordinaires

Les témoins doivent être préparés exactement de la même façon que les concentrations d'essai. Il faut choisir les organismes en même temps et suivant la même méthode. Les récipients doivent être du même type, et l'eau de dilution, le sédiment témoin ou tout autre substrat doivent être uniformes dans tous les récipients. Les témoins doivent être disposés au hasard parmi les autres récipients. Ce n'est que de cette façon que l'on peut attribuer sans biais un effet à une autre cause que le traitement ou la matière à l'étude.

Les témoins servent de points de comparaison, mais, dans ses méthodes, Environnement Canada insiste pour que le point de comparaison indique des conditions et des modes opératoires satisfaisants. Les exigences particulières concernant les performances du témoin varient selon le type d'essai, mais on peut donner des exemples. Dans l'essai sur la croissance et la survie de vers polychètes, le taux moyen de survie, chez les témoins de ces vers, doit être d'au moins 90 % (EC, 2001a). Pour ce qui concerne la croissance des larves de têtes-de-boule, pas plus de 20 % des larves témoins peuvent être moribondes ou présenter un comportement de nage atypique à la fin de l'essai et elles doivent également atteindre le poids sec moyen de 250 µg (EC, 1992b). Chez les embryons de salmonidés (EC, 1998a), le pourcentage moyen d'œufs témoins non viables ne doit pas être excéder 30 %. Chez les témoins de lentille d'eau, le nombre de frondes doit avoir au moins octuplé (EC, 1999b).

2.7.2 Témoins du solvant

Parfois une substance dont on étudie la toxicité est faiblement soluble dans l'eau. Un solvant pourrait aider à obtenir les fortes concentrations dont on a besoin pour observer un effet puissant. Habituellement, cela s'appliquerait aux essais de toxicité en milieu aquatique (pour le poisson, les algues, etc.) ou aux essais sur un sédiment que l'on enrichirait en la substance en question. Cela pourrait également s'appliquer à un essai de toxicité avec un sol, si la substance était ajoutée au sol sous forme de solution.

Dans les essais de toxicité d'Environnement Canada, on préfère n'utiliser que de l'eau de dilution comme diluant de la substance à l'étude ; on devrait éviter tout autre solvant, sauf nécessité absolue (EC, 1997a, b ; 2001a). Si on a besoin d'aide pour diluer une substance faiblement soluble, le premier choix est une colonne génératrice (Billington *et al.*, 1988). La dispersion ultrasonique serait moins conseillée, tandis que les solvants organiques, les émulsifiants ou les dispersants le seraient encore moins (EC 1997b ; 1998a ; 2001a). Parfois, les conditions à respecter sont plus

rigoureuses : aucun solvant ne devrait être utilisé dans l'essai, sauf s'il entre dans composition de la préparation de la substance normalement vendue dans le commerce (EC, 1992f ; 1998a ; 1999b).

Le plan d'expérience d'un essai d'Environnement Canada utilisant un solvant doit prévoir un *témoin du solvant,* c'est-à-dire une enceinte expérimentale (ou des enceintes répétées) en tout point semblables au témoin ordinaire, sauf qu'elle(s) renferme(nt) du solvant à la concentration maximale employée dans l'essai. Ce témoin s'ajoute au témoin habituel. Inutile de dire que la concentration du solvant devrait être très inférieure à sa concentration toxique et, autre condition à respecter parfois, elle ne doit pas excéder 0,1 mL/L (EC, 1992f ; 1999b). Si sa toxicité est inconnue, on devrait l'estimer de la façon habituelle afin de déterminer son seuil d'effet, avant de l'utiliser dans tout autre essai (EC 1997a, b ; 1999b).

Les effets du témoin du solvant ne doivent pas être plus puissants que ceux du témoin ordinaire. Telle est l'exigence formulée dans certaines méthodes d'Environnement Canada, qui ne précisent pas, à cet égard, de méthodes statistiques précises (EC, 1992f ; 1998a ; 1999b). Dans certaines méthodes (EC, 1998a ; 1999b), si on prévoit un témoin du solvant, celui-ci devient automatiquement le témoin qui servira de façon générale à évaluer l'effet du toxique. Cependant, dans les essais sur un sédiment employant des larves de chironomes, *H. azteca,* des amphipodes marins et des vers polychètes, le témoin du solvant est uniquement utilisé de cette façon si le paramètre de sa toxicité diffère statistiquement de celui du témoin ordinaire (EC, 1992e ; 1997b ; 2001a). Il n'est pas souhaitable de grouper les résultats donnés par le témoin du solvant avec ceux que donne le témoin ordinaire, et cette interdiction est absolue dans l'essai employant des truitelles (EC, 1998a), parce que, dans l'eau témoin ou de dilution, il manque un facteur (le solvant) qui pourrait agir sur les organismes aux autres concentrations. Bien que, dans l'essai sur un sédiment employant des polychètes (EC, 2001a), les données provenant du témoin du solvant doivent être réunies à celles qui proviennent du témoin ordinaire si les deux ne sont pas différents selon le test *t,* on peut très bien se passer de cette opération. L'OCDE (OECD, 2004) ne favorise pas un tel groupage et elle fait remarquer que des différences réelles existant entre les deux témoins pourraient avoir échappé au test statistique.

Dans toute méthode d'essai, les organismes soumis au témoin du solvant doivent satisfaire aux critères de validité de l'essai normalement applicables au témoin.

2.7.3 Témoins de la salinité

Un témoin de la salinité est une enceinte témoin séparée ou un ensemble d'enceintes destinées à évaluer l'effet de salinités moins qu'optimales dans un essai de toxicité employant des organismes marins. Ce témoin sert aussi de témoin normal. On pourrait devoir utiliser des témoins de la salinité dans des essais employant un organisme marin, que ce soit dans des milieux tels que l'eau ou un sédiment.

Essais avec salinités non ajustées. — Un témoin de la salinité serait souhaitable dans un essai sans ajustement des salinités. Par exemple, on pourrait vouloir évaluer l'effet total d'un effluent constitué d'eau douce et rejeté en milieu marin. On devrait alors utiliser des témoins de la salinité, outre le témoin de l'eau témoin ou de dilution, à une salinité favorable (30 ‰). Dans ces récipients supplémentaires, les salinités devraient être identiques à celles des enceintes expérimentales ou couvrir le même intervalle. Environnement Canada précise que les témoins de la salinité devraient être préparés par ajout d'eau distillée ou désionisée à l'eau salée témoin ou de dilution se trouvant dans une série de récipients de sorte que les concentrations seraient celles du liquide dont on étudie la toxicité (EC, 1992f). Le même mode opératoire serait logique si on étudiait la toxicité d'un sédiment (disons de déblais de dragage) destiné à l'immersion en mer, qui renfermerait une fraction liquide constituée essentiellement d'eau douce.

Visiblement, le but des témoins de la salinité est de révéler tout effet nuisible d'une faible salinité agissant seule. Ces témoins ne révéleraient cependant pas l'effet nocif aggravé par l'action combinée d'un écart de salinité et de la matière à l'étude. Pour interpréter les résultats, on ne peut qu'attribuer à cette matière toute toxicité supérieure à celle que l'on aura constatée chez les témoins de la salinité.

Dans un essai employant des épinoches (EC, 1990b), on n'ajusterait normalement pas la salinité. Un témoin de la salinité n'est pas exigé, mais il serait avantageux. La méthode offre la possibilité d'ajuster la salinité à

28 ‰, pour l'essai d'un produit chimique, d'un effluent, d'un percolat et d'un élutriat.

En théorie, il pourrait arriver que la salinité défavorable ait été excessive. Un effluent pourrait être très salé, et on pourrait le soupçonner de renfermer des matières toxiques. Les principes présidant à l'utilisation de témoins de la salinité et à l'interprétation des résultats qu'ils permettraient d'obtenir resteraient les mêmes que dans le cas d'une faible salinité.

Essais dans lesquels on ajuste la salinité. — Dans les essais de toxicité en milieu marin, l'usage est d'ajuster toutes les concentrations à une seule salinité favorable. C'est ainsi que l'on procède habituellement avec les oursins (EC, 1992f) ainsi que dans les essais sur un sédiment employant des amphipodes (EC, 1992e). On procède toujours ainsi dans le cadre du programme d'Environnement Canada de Suivi des effets sur l'environnement, dans les quatre essais employant des organismes marins (EC, 2001b). Pour cet ajustement, Environnement Canada a adopté la salinité favorable normale de 30 ‰.

Dans ces essais, il n'y aurait pas de témoin de la salinité. Il y aurait un témoin normal, d'une salinité de 30 ‰, préparé avec la même matière que celle qui a servi à ajuster la salinité des concentrations d'essai et (ou) de l'eau de dilution.

Témoins particuliers de la salinité. — Dans les essais en milieu marin dans le cadre du Programme de suivi des effets sur l'environnement (EC, 2001b), on pourrait avoir besoin d'un autre type de témoin de la salinité. Cela concerne la technique particulière utilisée pour l'ajustement de la salinité.

On peut augmenter la salinité d'un effluent ou d'une « concentration d'essai » par l'ajout de sels secs (de qualité « réactif » ou un mélange du commerce) ou d'une saumure sursalée. Normalement, on préparerait toutes les concentrations expérimentales et tous les témoins avec la même matière, auquel cas aucun témoin particulier supplémentaire ne serait nécessaire. [On pourrait préciser le mode de préparation (avec des sels secs ou de la saumure sursalée, selon le cas) des témoins normaux (EC, 2001b).]

Si, cependant, l'eau de dilution utilisée pour la préparation des concentrations d'essai a une origine différente de celle du ou des témoins préparés avec des sels secs ou une saumure sursalée, on devrait préparer un second témoin ou ensemble de témoins avec de l'eau de dilution (témoins d'eau de dilution)[12]. La salinité de tous ces traitements seraient de 30 ‰.

Analyses statistiques des témoins de la salinité. — Le principe présidant à l'interprétation des témoins est le suivant : chaque type de témoin, individuellement, doit satisfaire aux conditions de performance spécifiées pour le témoin dans les instructions relatives à l'essai particulier de toxicité. Par exemple, dans un essai réalisé dans le cadre du Programme canadien d'étude de suivi des effets sur l'environnement, le témoin préparé avec des sels secs devrait satisfaire aux critères spécifiés, tout comme devrait le faire le témoin préparé avec de l'eau de dilution, si on a utilisé les deux types de témoins. Si une catégorie, n'importe laquelle, de témoins ne satisfaisait pas aux exigences, on considérerait l'essai de toxicité comme invalide.

Cet échec est le plus susceptible de survenir dans un essai dont on n'aurait pas normalisé la salinité des diverses concentrations d'essai. Une forte concentration d'effluent d'eau douce abaisserait la salinité dans l'enceinte expérimentale. Dans ce cas, les témoins correspondants de la salinité pourraient ne pas satisfaire aux normes de performances. La conclusion serait évidente : tous les effets constatés aux fortes concentrations d'essai seraient probablement attribuables, en tout ou en partie, à la faible salinité. Ce ne serait pas un essai valide pour déterminer les effets du toxique (l'effluent).

Si tous les types de témoins avaient satisfait aux exigences de performances, l'essai de toxicité serait valide. L'emploi des résultats dus au(x) témoin(s) dans l'analyse des constatations suivrait alors n'importe quelle pratique normale spécifiée dans la méthode

12. L'eau de dilution pourrait être de l'eau de mer non contaminée, de 30 ‰de salinité, alors que l'on aurait employé des sels secs pour ajuster à la même valeur la salinité de l'effluent à l'étude. Par ailleurs, l'eau de dilution pourrait être préparée avec de la saumure sursalée et de l'eau désionisée, tandis que la salinité de l'effluent aurait été ajustée au moyen de sels secs. D'autres combinaisons sont possibles. Le principe à respecter est que des témoins particuliers sont nécessaires si la préparation des témoins diffère en quoi que ce soit de celle des concentrations expérimentales.

particulière d'essai.

2.7.4 Sédiments et sols témoins et de référence

Dans des essais de toxicité d'un sédiment ou d'un sol d'Environnement Canada, le mode opératoire normalisé prévoit l'emploi d'un *sédiment ou d'un sol de référence* avec chaque échantillon ou ensemble d'échantillons provenant d'un endroit donné (par ex. EC, 1997a). On présume que l'échantillon de référence n'est pas pollué et on compare les résultats que donne le ou les échantillons aux résultats donnés par la matière de référence pour déceler tout effet tel qu'un taux accru de mortalité ou une taille plus petite. En conséquence, l'échantillon de référence sert d'étalon à l'essai. Cela est logique, parce que cela permet une évaluation localisée de la toxicité.

Avec chaque lot d'échantillons, les essais utilisent également un *sédiment ou un sol témoin* pour vérifier la qualité générale de l'essai et des organismes qui y sont employés. Une limite de performance acceptable est fixée dans chaque méthode d'Environnement Canada. Par exemple, il ne peut pas y avoir plus de 30 % de mortalité dans l'essai sur un sédiment employant des chironomes (EC, 1997a). Le sédiment ou le sol témoin ne sert normalement pas de base de comparaison directe des effets observés dans les échantillons. Cependant, il serait utilisé à cette fin si le sédiment de référence se révélait peu convenable à la comparaison, en raison de sa toxicité ou de caractéristiques physicochimiques atypiques (EC, 1997a). L'approche est raisonnable.

Les deux types de témoins sont définis dans le glossaire, mais nous pourrions en distinguer ici les caractéristiques. On prélève un *sédiment de référence* sur le terrain, dans le voisinage général des stations d'étude, dans un emplacement que l'on estime être à l'abri de l'influence de la source de contamination à l'étude. On présume que ce sédiment de référence n'est pas pollué et qu'il possède des caractéristiques physiques presque identiques à celles des échantillons à l'étude. Le *sol de référence* est prélevé en milieu terrestre, mais, par ailleurs, il possède les caractéristiques et les fonctions du sédiment de référence. Parce qu'il constitue le témoin, il intègre dans l'essai les *effets de matrice.* Il peut aussi servir de diluant pour la préparation de dilutions du sol à l'étude.

Le *sédiment ou le sol témoin* ne serait pas prélevé dans le même voisinage en général que les échantillons. On pourrait le prélever dans un lieu non contaminé ou le préparer avec les constituants appropriés. On veut obtenir un sédiment ou un sol non contaminé dans lequel on sait que les organismes prospèrent. Ce pourrait être le substrat d'où les organismes ont été prélevés ou dans lequel ils ont été élevés ou cultivés.

Analyses statistiques. — On compare l'effet à celui du sédiment ou du sol de référence à moins que cela ne convienne pas, auquel cas on le remplace par le sédiment ou le sol témoin. L'analyse et l'interprétation suivent les méthodes normalisées décrites dans d'autres parties du document et elles sont exemplifiées dans le document portant sur les vers polychètes (EC, 2001a). Les essais à concentration unique sont limités au test d'hypothèse (section 7). Si les essais emploient des dilutions de la matière à l'étude ou des sédiments ou des sols enrichis, les analyses peuvent aboutir à des estimations ponctuelles, de la CI p (section 6) ou de la CE 50 (section 4).

2.8 Toxiques de référence et cartes de contrôle

Repères

- *Les essais réalisés périodiquement avec un toxique étalon (de référence) visent à évaluer les variations de sensibilité des organismes et la précision intralaboratoire.*
- *Les résultats successifs qu'obtient un laboratoire sont portés sur une carte de contrôle. La nouvelle valeur est comparée à la moyenne des résultats antérieurs et à la zone de confiance de ± 2 écarts types. Tous les calculs se fondent sur des concentrations logarithmiques. Les laboratoires canadiens oublient souvent de le faire.*

Les essais avec un *toxique de référence* sont tout à fait différents dans leur but et leurs caractéristiques des témoins décrits dans le § 2.7. Ils utilisent un toxique étalon titré pour mesurer les effets relatifs subis par les organismes en expérience et ils sont normalement répétés au cours des mois où le laboratoire fonctionne. Ces essais sont destinés à : *a*) déceler toute

modification de la sensibilité des organismes au fil du temps ; *b*) évaluer toute fluctuation de la technique de mesure utilisée par le laboratoire. On ne peut pas distinguer ces deux causes de variation, particulièrement parce qu'il n'est pas usuel de répéter les essais. L'essai est tout à fait distinct de tout essai de toxicité sur des échantillons, bien qu'il ait souvent lieu simultanément. Les méthodes d'essai d'Environnement Canada employant des organismes aquatiques et terrestres exigent l'emploi périodique de toxiques de référence.

Le toxique de référence peut également être synonyme de *témoin positif*. Les toxiques de référence communément utilisés sont le phénol, le chlorure de sodium ou un métal (EC, 1999a). Les résultats doivent être portés sur une carte de contrôle, pour juger si la variation des résultats au laboratoire est satisfaisante. La carte de contrôle pourrait être semblable à celle de la figure 2, bien que le tracé des points et le légendage puissent être faits à la main.

La fig. 2 montre les résultats d'une partie d'une série d'essais effectués avec un toxique de référence et ayant employé la truite arc-en-ciel. Nous avons modifié l'échelle de temps et les dates fournies par un laboratoire canadien. On constate que trois essais ont eu lieu dans chaque trimestre, c'est-à-dire un essai mensuellement. La moyenne du logarithme des CE 50 est – 0,027356. L'antilogarithme, c'est-à-dire 0,94 mg/L, est la moyenne géométrique des CE 50, ce que montre une ligne traversant le graphique. La zone de confiance, calculée comme étant ± 2 écarts types, est délimitée par les lignes horizontales. Elle sert d'indicateur visuel de la dispersion des résultats. L'écart type calculé d'après les données de la fig. 2 est de 0,15288, de sorte que 2 écarts types égaleraient 0,30576. L'addition et la soustraction de cette valeur à (de) la moyenne donnent les logarithmes de – 0,33312 et de 0,27840 comme limites de la zone de confiance, soit les antilogarithmes de 0,46 et de 1,9 mg/L (représentés dans fig. 2).

Quand un laboratoire obtient une nouvelle CE 50 pour le toxique de référence, il porte cette valeur sur la carte de contrôle comme dans la fig. 2. Si cette valeur se situe dans la zone de confiance, on la considère comme satisfaisante. Si elle se situe à l'extérieur de la zone de confiance, le laboratoire devrait rechercher les causes de cet écart. La nouvelle CE 50 serait alors intégrée dans les résultats de tous les essais antérieurs effectués au laboratoire avec le toxique de référence, et on recalculerait la moyenne géométrique et les limites de la zone de confiance. Ces limites s'appliqueraient au prochain essai du toxique de référence. Il n'est pas difficile de programmer un tableur pour bien effectuer de tels calculs séquentiels et produire un graphique semblable à celui de la fig. 2.

Tous les calculs destinés à la carte de contrôle doivent employer les logarithmes de la concentration, et l'axe vertical de la fig. 2 emploie une telle échelle. Dans le § 2.3, on explique pourquoi il faut employer les logarithmes, tandis que le mode de calcul de la moyenne et de l'écart type sont présentés dans l'annexe F. Si on avait utilisé les valeurs arithmétiques des CE 50, en effectuant dans tous les cas des calculs arithmétiques, on aurait obtenu une zone de confiance différente. La moyenne aurait été 0,99 mg/L, ce qui est quelque peu supérieur à la valeur correcte de 0,94 mg/L. La limite supérieure de la zone de confiance aurait été 1,6 au lieu de 1,9 mg/L, tandis que la limite inférieure aurait été de 0,39 au lieu de 0,46 mg/L. L'intervalle arithmétique entre ces limites aurait été plus étroit, de 1,2 au lieu de 1,4 mg/L (annexe F).

Dans des laboratoires canadiens, l'une des carences méthodologiques les plus répandues est la résistance des expérimentateurs au calcul logarithmique. En effet, beaucoup de méthodes antérieures publiées par Environnement Canada en faisaient une recommandation et non une obligation.

Des progiciels offrent la possibilité de calculer et de tracer une carte de contrôle. Les expérimentateurs devraient vérifier que les calculs emploient les logarithmes de la concentration. Par exemple, le progiciel CETIS (2001) tracera la carte avec la moyenne ± 2 écarts types, mais en calculant de façon erronée les valeurs arithmétiques des concentrations. La même erreur est faite par le progiciel TOXCALC (1994).

2.8.1 Variation raisonnable

L'étendue de la zone de confiance est visiblement importante. Si elle est étroite, cela signifie que le laboratoire a obtenu des résultats d'une grande précision. En conséquence, si une CE 50 tombe à

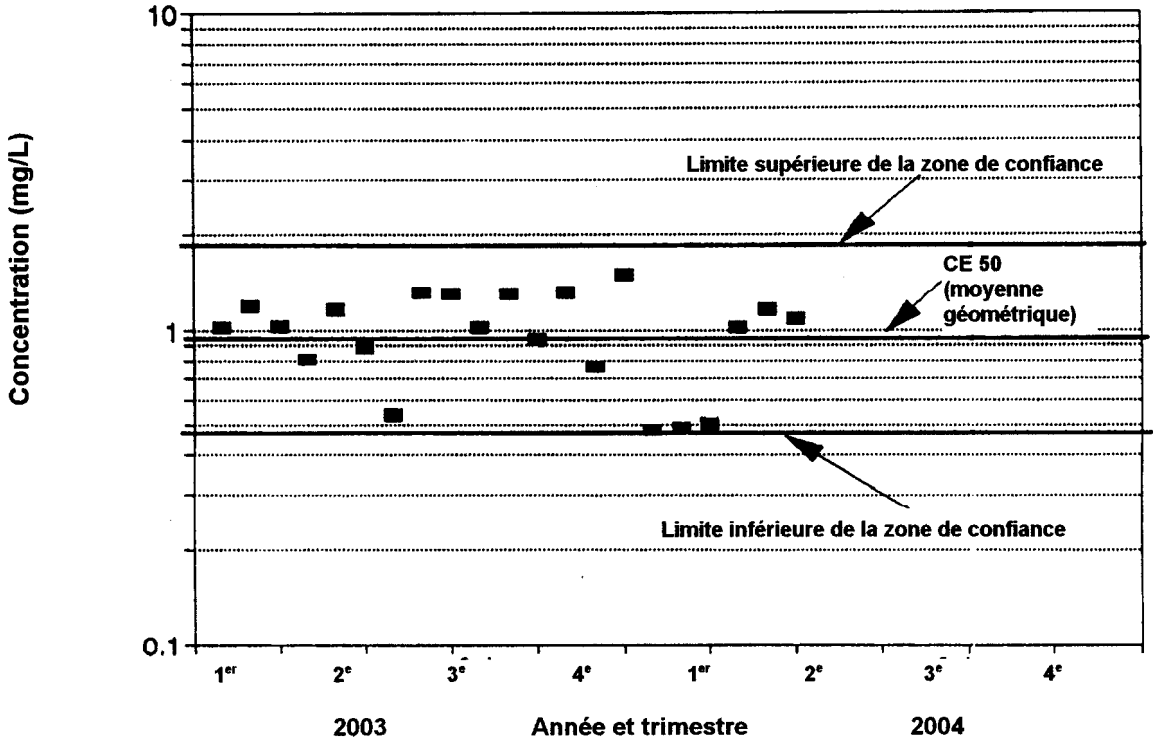


Figure 2. — Carte de contrôle pour les essais avec un toxique de référence. Cette carte montre les résultats réels obtenus par un laboratoire canadien, dans des essais d'un toxique de référence en milieu aquatique. Les CE 50 se maintiennent assez régulièrement dans la zone de confiance (certaines y entrent de justesse). Globalement, la variation est légèrement plus grande que ce que l'on pourrait considérer comme souhaitable. L'axe vertical et tous les calculs se fondent sur les valeurs logarithmiques des CE 50.

l'extérieur, cela ne signifie pas nécessairement un problème grave de mode opératoire ni une variation importante de la sensibilité des organismes. De fait, environ 5 % des CE 50 devraient tomber à l'extérieur de la zone de ± 2 écarts types par le seul fait du hasard. Inversement, un laboratoire pourrait avoir eu des résultats erratiques qui se matérialiseraient par une zone de confiance large ; les CE 50 ultérieures pourraient se trouver à l'intérieur, mais elles indiqueraient néanmoins une variation indésirable.

En conséquence, on peut envisager une deuxième façon d'évaluer la variation, que l'on pourrait appeler « degré raisonnable de variation ». Ce sujet diffère de celui de la zone de confiance que nous venons de décrire. Environnement Canada n'a pas formellement défini le degré raisonnable de variation auquel on pourrait s'attendre d'un ensemble d'essais répétés. Cependant, il a été proposé qu'un coefficient de variation (C.V.) de pas plus de 30 % et, de préférence, d'au plus 20 % pourrait être raisonnable pour les essais avec des toxiques de référence (EC, 1990d)[13]. Pour la variation des toxiques de référence, on a proposé le même nombre-guide dans la méthode d'essai d'un sédiment avec des vers polychètes (EC, 2001a).

Le nombre-guide préconisé par Environnement Canada se fondait apparemment sur des moyennes et écarts

13. Le coefficient de variation (C. V.), habituellement exprimé en pourcentage, égale l'écart type divisé par la moyenne, formule valable pour les données sous forme arithmétique. En conséquence, connaissant le C. V., on peut calculer l'écart type : on multiplie simplement la moyenne par le C. V. sous forme de fraction décimale. On ne peut pas calculer de la sorte le C. V. avec des données logarithmiques. Dans le cas des données log-normales, la formule est comme suit : C. V. = racine carrée de $[10^{(s \times s)} - 1]$, où s est l'écart type calculé à partir des données logarithmiques. L'écart type au carré est la variance, qui, dans la formule, pourrait remplacer l'exposant.

types arithmétiques d'ensembles de CE 50 et, de ce fait, il comporte un certain biais. Cependant, on peut calculer un nombre-guide équivalent pour le logarithme des concentrations, ce que nous avons fait dans le présent document (v. l'annexe F). Une variation raisonnable d'après ce nombre-guide ou le pifomètre serait comme suit : **la valeur de l'écart type calculé avec des données logarithmiques ne devrait pas excéder 0,132 et, de préférence, ne devrait pas excéder 0,0338.** Ces mêmes logarithmes peuvent servir à estimer l'écart type de tout *ensemble de paramètres de toxicité*. Ils correspondent approximativement aux C. V. arithmétiques précédemment mentionnés de 30 et de 20 %, mais ils évitent la distorsion possible.

Le nombre-guide préconisé ici exige le calcul de la moyenne et de l'écart type de l'ensemble des paramètres de toxicité au moyen de logarithmes de concentration . On compare l'écart type calculé (un logarithme) à 0,132, et s'il est égal ou inférieur, on considère comme acceptable (« raisonnable ») la variation de l'ensemble de paramètres de toxicité.

On pourrait comparer le nombre-guide d'une variation acceptable aux données de la fig. 2. Ces données ont un écart type logarithmique calculé de 0,15288..., de sorte que la variation observée dans les paramètres de toxicité excède quelque peu la variation « raisonnable » selon notre nombre-guide.

On peut se faire une idée de cette variabilité « raisonnable » en imaginant un scénario hypothétique pour la comparer à la situation décrite dans la fig. 2. Si l'ensemble hypothétique de CE 50 avait la même moyenne, mais se trouvait à avoir l'écart type « raisonnable » de 0,132, les limites de la zone de confiance seraient de 0,51 et de 1,7 mg/L (v. l'annexe F) [14]. La zone de confiance pour les données hypothétiques « raisonnables » serait quelque peu plus étroite que celle de la fig. 2.

Si la variation d'un ensemble de données hypothétiques était même moindre, avec un écart type égal à la valeur « préférable » de 0,0338 et que la moyenne était égale à celle de la fig. 2, la zone de confiance serait très étroite. Converties en valeurs arithmétiques pour les besoins de la comparaison avec la fig. 2, les limites seraient 0,80 et de 1,1 mg/L, ce qui semble quelque peu optimiste pour la variation entre des essais répétés de toxicité.

Il n'y aurait pas de relation constante entre cette proposition pour évaluer une variation raisonnable et les limites de longue date de la zone de confiance selon Environnement Canada, qui sont de ± 2 écarts types pour les toxiques de référence. La règle empirique (ou le nombre-guide) de la variation « raisonnable » resterait constante, mais les limites de la zone de confiance varieraient selon l'ensemble de données.

2.9 Transformation des données sur l'effet

Repères

- *Pour ce qui concerne les données quantiques, une transformation ordinaire et standard utilise le probit ou le logit de l'effet pour estimer la CE 50.*
- *Dans l'estimation des paramètres de toxicité sublétale par régression, on pose, par hypothèse, que les résidus suivent la loi normale et que la transformation peut aider à cela. On peut également simplifier la relation pour l'utilisation de la régression. Un inconvénient majeur est que la transformation exige une pondération individualisée pour compenser la modification des variances des groupes d'observations. Cela exige des conseils ou de la compétence en statistique.*
- *Dans le test d'hypothèse, si les données relatives aux effets ne satisfont pas aux exigences en matière de normalité et d'homogénéité de la variance, une transformation pourrait y remédier et permettre l'analyse suivant les méthodes paramétriques ordinaires. Cela est recommandé comme première option, si on veut effectuer un test d'hypothèse et si les données ne satisfont pas aux exigences.*
- *On utilise généralement pour les transformations les logarithmes et les racines carrées. La transformation arc sinus racine carrée est réservée aux données quantiques ; cette transformation et sa réciproque ne sont pas souvent utiles.*

14. Les expérimentateurs ayant perdu l'habitude des calculs logarithmiques pourraient vérifier leurs modes opératoires en consultant le glossaire, le § 2.3.5, l'annexe D ou en consultant le traitement arithmétique de l'annexe F.

La transformation de résultats pourrait aider à l'une ou l'autre des deux approches générales d'analyse des résultats d'essais d'écotoxicité — les techniques de régression et le test d'hypothèse(s). Les deux approches sont assorties de certaines exigences concernant la distribution normale des données sur les effets. Si ces données ne suivent pas la loi normale, il est possible de les transformer pour qu'elles satisfassent à cette exigence. Dans le cas de la régression, la transformation peut devoir en outre épouser la forme d'une droite afin de simplifier l'analyse.

2.9.1 Utilisation en régression

Pour les essais de létalité aiguë ou d'autres **effets quantiques,** il est usuel de transformer les pourcentages d'effet en probits ou en logits. Ces transformations sont appropriées et avantageuses pour l'analyse statistique. En général, les probits ou les logits redressent la courbe sigmoïde d'un ensemble de données (annexe H), ce qui autorise un modèle de droite et diminue le nombre de paramètres à estimer. Ces avantages sont décrits dans le texte qui suit sous la rubrique **« Avantage de la transformation... »** L'utilisation classique des probits ou d'autres transformations pour les données quantiques est l'objet d'une discussion supplémentaire dans la section 4. (Pour ce qui concerne les concentrations, on conserve les logarithmes pour l'analyse [15].)

Les analyses des résultats des essais de toxicité sublétale adoptent des techniques plus avancées, notamment la régression non linéaire (§ 6.5). La construction d'intervalles de confiance pour les paramètres des modèles de régression non linéaire pose d'habitude comme hypothèse la distribution normale des résidus. De nouveau, la transformation serait une approche possible pour satisfaire à cette exigence.

Avantage de transformation : la simplicité. — Un principe important des techniques de régression appliquées aux estimations ponctuelles est de maintenir la simplicité du modèle, si cela peut se faire raisonnablement. La transformation des données peut simplifier la relation et permettre l'emploi d'un modèle simple. Bien que l'on puisse créer des modèles pour s'ajuster à une relation complexe, l'équation résultante possédera de nombreux termes et, par conséquent, fera perdre des degrés de liberté, affaiblira le pouvoir prédictif et, peut-être, élargira la zone de confiance du paramètre prévu de toxicité (Andersen *et al.*, 1998). Les statisticiens insistent sur cette qualité, par ex. : « Ainsi la simplicité, matérialisée par la parcimonie des paramètres, est... une qualité de tout modèle... Non seulement un modèle parcimonieux permet-il au chercheur ou à l'analyste de réfléchir à ses données, mais un modèle juste en grande partie permet de meilleures prévisions qu'un modèle qui s'embarrasse de paramètres inutiles » (McCullagh et Nelder, 1989). Il s'ensuit que les transformations des données pourraient être avantageuses en permettant de simplifier le modèle.

L'équation ajustée à la croissance exponentielle est un simple exemple de transformation (v. le § 6.5.3).

$$Y = \alpha\beta^X \qquad \Rightarrow$$

$$\log Y = \log \alpha + X \log \beta \qquad [1]$$

Grâce aux logarithmes, on peut transformer cette équation, qui est une relation multiplicative (1re ligne de l'équation 1) en une relation linéaire (seconde ligne de l'équation 1), ce qui donne lieu à une régression relativement simple.

Une transformation fréquente, qui permet d'ajuster les données proportionnelles à une distribution normale avec équivariance consiste à prendre l'arc sinus des effets (v. le glossaire et le § 2.9.3). Ce type de transformation semble une façon facile de simplifier l'analyse et d'aplanir l'« obstacle que constitue le calcul d'intervalles de confiance autour d'estimations par régression non linéaire... » (Nyholm *et al.*, 1992).

Inconvénients pour la régression. — Des complications très graves tendent à neutraliser l'avantage apparent de la transformation. Bien que celle-ci puisse viser à simplifier l'estimation de paramètres, elle peut déformer une véritable relation (mécaniste). Par exemple, les réactions enzymatiques sont décrites de façon mécaniste par l'équation non linéaire de Michaelis-Menten. De véritables effets de

15. L'utilisation de valeurs logarithmiques de la concentration (et/ou du temps) dans l'analyse permet simplement de conserver les unités originales de l'exposition, pour des motifs scientifiques fondamentaux (§ 2.3). La transformation des logarithmes de la concentration en valeurs arithmétiques pour un ensemble de calculs, outre qu'elle est erronée, introduirait probablement une asymétrie dans la relation et exigerait un modèle plus complexe.

concentration de seuil pourraient également être déformés par des transformations inopportunes.

Parfois, la transformation peut mener à des estimations fortement biaisées des paramètres de toxicité, décrites à l'occasion comme « fatales ». Ces problèmes sont discutés du point de vue toxicologique par Christensen et Nyholm (1984) et par Nyholm *et al.* (1992). Ces auteurs font remarquer que la transformation a besoin d'une pondération appropriée pour compenser l'altération, à différents degrés, des variances des points de données. La pondération est propre aux données obtenues, de sorte qu'il n'existe pas de progiciel statistique applicable à la façon d'un livre de cuisine. Les coefficients de pondération doivent être inversement proportionnels à la variance des données calculée pour les mesures originales (observations) à n'importe quelle valeur donnée de la variable indépendante X (ou d'habitude, log X). Même alors, la compensation grâce à la pondération pourrait ne pas être suffisamment précise pour les données irrégulières ou pour les observations près des extrêmes de la distribution de la relation dose-effet. La pondération devrait aussi tenir compte de l'expression de la variabilité originale en unités absolues ou proportionnelles à la grandeur de la variable mesurée.

Cette personnalisation statistique des ensembles de données excède amplement les limites que l'on peut définir dans les méthodes d'Environnement Canada relatives aux essais en routine. Les expérimentateurs devraient être conscients des pièges dans lesquels ils risquent de donner s'ils transforment les résultats pour obtenir une régression linéaire. En outre, il leur est conseillé, s'ils songent que la transformation pourrait être utile, de consulter un statisticien qui connaît bien les essais de toxicité. Il se peut qu'il existe déjà des méthodes statistiques convenables (v. le § 6.5.8) ou que des progiciels de statistique deviennent disponibles.

2.9.2 Utilisation pour le test d'hypothèse

Les méthodes les mieux connus de test d'hypothèse supposent que les résultats suivent une distribution normale. C'est la supposition retenue pour les tests *t*, l'analyse de variance et les tests de comparaisons multiples. En conséquence, il faut tester la normalité des données avant d'en faire l'analyse (§ 7.3).

Si un ensemble de données ne suit visiblement pas la distribution normale, l'expérimentateur a le choix entre trois partis principaux :

- utiliser une méthode paramétrique plus sophistiquée, adaptée aux données ;
- transformer les données pour rendre la distribution normale ;
- utiliser une méthode non paramétrique qui ne formule aucune hypothèse sur la distribution.

Le premier parti est le plus souhaitable, mais on le choisit rarement parce que la plupart des expérimentateurs, qui ne sont pas statisticiens, n'en connaissent pas les méthodes. Dans leur évolution, les méthodes usuelles ont fini par se fonder sur d'autres approches, parce que les méthodes paramétriques plus sophistiquées comportaient des calculs difficiles, mais cela n'est plus un obstacle depuis l'avènement des ordinateurs. Nous ne donnons pas de conseils sur ces méthodes plus avancées, mais nous les évoquons dans les § 6.5.2 et 6.5.11 sur les modèles linéaires généralisés (GLIM). Il faut espérer que les échanges avec les statisticiens rendront disponibles et adaptables à l'écotoxicologie de telles méthodes améliorées.

On a préféré le deuxième parti (transformer les données pour rendre la distribution normale). Cette utilisation de la transformation visait à obtenir des données appropriées aux méthodes statistiques des décennies antérieures. Elle permet l'emploi de méthodes bien connues d'analyse, dont les marches à suivre sont relativement simples et dont les tables statistiques sont facilement accessibles.

Le troisième parti énuméré (l'emploi d'une méthode non paramétrique d'analyse) est également devenu une approche moderne usuelle, en partie grâce au développement et à la programmation de méthodes usuelles pour le test d'hypothèse aux États-Unis. On a habituellement eu recours aux méthodes non paramétriques quand l'analyse paramétrique n'est pas valable. Dans des nombreux cas, elles sont moins puissantes que les tests paramétriques correspondants, pour ce qui est de distinguer les effets. À l'instar des tests paramétriques, les non paramétriques formulent des hypothèses sur les données, par ex. sur l'indépendance des observations et l'homogénéité de la variance, mais ils sont généralement plus robustes pour ce qui concerne les écarts par rapport à ces hypothèses.

Avantages et inconvénients. — Si l'on veut soumettre les résultats à un test d'hypothèse, la transformation peut accepter des mesures qui s'écartent de la normalité ou de l'homogénéité de la variance et les modifier en variables remplissant les conditions de l'analyse par des tests paramétriques connus. On peut aussi transformer certains ensembles de données quantiques pour mieux les plier aux tests d'hypothèses (§ 2.9.3).

En conséquence, on recommande une transformation appropriée, si nécessaire, comme option préférée pour les données qui ne satisfont pas aux exigences de la normalité et de l'homogénéité de la variance. Il est conseillé de consulter un statisticien. Le problème le plus grave est que l'on peut s'attendre à ce que la transformation modifie les relations entre les données. Il faut tenir compte des avertissements du § 2.9.1.

Si l'on n'a pas trouvé de transformation satisfaisante, le parti à prendre pourrait être l'analyse par des tests non paramétriques.

2.9.3 Transformations particulières

Les transformations des mesures les plus fréquemment utilisées sont le logarithme et la racine carrée. Les deux peuvent être efficaces si la variance augmente avec la moyenne. Les **logarithmes** sont utiles si l'effet tend à augmenter exponentiellement en raison de la concentration et si la variance est proportionnelle au carré du résultat moyen du traitement. Cela pourrait survenir avec la croissance de la population ou démographique ou le poids, et la transformation pourrait rendre la variance indépendante de la moyenne. La formule préférée, particulièrement si certaines des valeurs sont petites ou nulles, est log ($X + 1$).

La **racine carrée** peut également aider à stabiliser la variance. On peut également l'appliquer quand les données se présentent comme une série de dénombrements (distribution de Poisson), et que les variances des groupes sont proportionnelles aux moyennes. De nouveau, la formule préférée comprend une constante plutôt qu'une simple transformation, généralement la racine carrée de ($X + 0{,}5$), où X est une mesure individuelle (Zar, 1999). Lui est peut-être supérieure la transformation légèrement plus complexe de la racine carrée de X plus la racine carrée de ($X + 1$).

La **transformation réciproque** est rarement utile pour les données quantitatives. On ne recommande pas la transformation **arc sin racine carrée**, qui ne convient pas aux données quantitatives, parce qu'elle est destinée aux observations binomiales telles que les pourcentages ou les proportions (Zar, 1999). Parfois, cependant, l'expérimentateur pourrait souhaiter analyser les données quantiques par un test d'hypothèse, ce à quoi la transformation arc sinus pourrait être utile et convenir. Nous en discutons dans le § 7.2.6, et l'application de l'arc sinus est discutée dans le glossaire.

Essais à concentration unique

Repères

- *Les essais de toxicité à concentration unique sont d'ordinaire utilisés dans les études d'évaluation des sédiments et des sols contaminés ou pour la surveillance des effluents. Leurs résultats peuvent servir à juger du respect des règlements, à l'aide d'un critère réglementaire fixe du type « réussite ou échec », sans analyse statistique.*
- *Les tests visant à conclure à un effet statistiquement significatif d'après les résultats des essais quantiques de toxicité à concentration unique (par ex. la mortalité) dépendent de la nature et du plan du programme d'étude. Pour un échantillon prélevé en un endroit, sans répétition (réitération), l'essai pourrait se faire par comparaison avec le groupe témoin à l'aide de la méthode exacte de Fisher ou des tables de Finney. Pour un emplacement unique avec échantillons réitérés, par ex. une étude des sédiments ou des sols contaminés, on pourrait soumettre les résultats à la méthode exacte de Fisher.*
- *Dans le cas d'une étude portant sur plusieurs emplacements, sans répétition, et portant sur des effets quantiques, les résultats ne seraient pas statistiquement testables. Avec des échantillons réitérés, on pourrait évaluer les résultats au moyen d'une régression logistique effectuée par un statisticien ou sous sa surveillance. Parfois, l'analyse de variance pourrait être réalisable.*
- *Les essais quantitatifs à concentration unique (par ex. effets de l'exposition à un sédiment contaminé sur le poids atteint par les organismes) emploient des méthodes statistiques différentes. Pour l'échantillonnage réitéré dans un emplacement, on pourrait comparer les résultats à ceux que donne le témoin avec un test* t. *Sans répétitions, les résultats ne seraient pas statistiquement testables.*
- *À l'égard des résultats quantitatifs portant sur plusieurs emplacements, on dispose d'un certain nombre d'approches. S'il n'y a pas de répétitions, on ne conseille aucune analyse statistique. Si on a des échantillons réitérés, l'analyse de variance serait une première étape, si les résultats s'y prêtent. Si l'hypothèse nulle d'une différence nulle a été rejetée, l'analyse peut se faire à l'aide de l'un des nombreux tests de comparaisons multiples. Pour les données ordonnées (gradient prévu), le test de Williams comparerait chaque emplacement avec le témoin. Dans le cas de données non ordonnées, le test de Dunnett les comparerait aux témoins, tandis que le test de Dunn-Sidak serait une seconde option. Pour une comparaison deux à deux (chaque emplacement avec chacun des autres), on recommande la méthode de la plus petite différence significative de Fisher, le test de Tukey étant une solution de rechange.*
- *Pour ce qui concerne les échantillons de terrain réitérés et les données quantitatives exigeant une analyse non paramétrique, il est recommandé, dans la plupart des cas, de vérifier l'hypothèse nulle avant de passer à un test de comparaisons multiples. Si les données sont ordonnées, on devrait comparer les emplacements avec le témoin à l'aide du test de Shirley. La comparaison deux à deux des données ordonnées débuterait par le test de Jonckheere-Terpstra, puis emploierait le test de Hayter-Stone si l'hypothèse nulle était rejetée. Dans le cas des données non ordonnées, la comparaison avec le témoin débuterait par le test de Fligner-Wolfe ou, si ce test n'est pas accessible, par le test de Kruskal-Wallis. Si on avait rejeté l'hypothèse nulle, on appliquerait le test de Nemenyi-Damico-Wolfe, avec, comme deuxième choix, le test de sommation des rangs de Wilcoxon et, comme remplacement possible, le test multiunivoque de Steel. Pour la comparaison deux à deux, le premier choix serait le test d'hypothèse de Kruskal-Wallis. Le test recommandé de comparaisons multiples est celui de Critchlow-Fligner-Steel-Dwass, avec comme solutions possibles de rechange, le test de comparaison par paires de Steel ou l'utilisation répétée du test de Kruskal-Wallis.*

Les essais à concentration unique sont souvent utilisés dans les programmes de surveillance de l'environnement pour contrôler la conformité des rejets

aux règlements et examiner des zones de sédiments, de sols ou d'eaux de surface potentiellement polluées. En dépit de leur manque de puissance, les essais sont une façon efficace et utile de jouer ce rôle d'examen.

Un programme de surveillance de la conformité des rejets liquides aux règlements pourrait simplement utiliser l'effluent tel quel, non dilué. Les effets de l'effluent seraient comparés à ceux du témoin le plus approprié que l'on pourrait choisir. Les essais initiaux portant sur un sol ou un sédiment potentiellement contaminé utilisent habituellement un échantillon non dilué. On en compare normalement les effets à ceux d'un sol ou d'un sédiment témoin et de référence (v. le glossaire) [16].

Aucune comparaison statistique avec le témoin n'est nécessaire quand on réalise un essai à concentration unique, tel que la mesure de la létalité aiguë, en vertu des règlements visant les fabriques de pâtes et papiers et les mines de métaux du Canada. La matière à l'étude échouerait ou réussirait (à l'essai) selon que la mortalité aurait excédé ou non la limite permise.

Il existe des méthodes statistiques toutes faites pour certains autres essais à concentration unique, pour l'analyse des résultats, et on utilise souvent le test d'hypothèse avec prélèvement réitéré, au besoin, d'échantillons sur le terrain. En analyse, il faut maintenir la distinction entre les résultats quantiques et les résultats quantitatifs.

Dans la fig. 3, on expose les plans courants d'expérience et les choix de méthodes d'analyse qui leur correspondent. Dans les paragraphes qui suivent, on discute de ces plans et méthodes. Les essais à concentration unique peuvent connaître de nombreuses variations pour répondre à une situation particulière. Pour ce qui concerne les plans spéciaux d'expérience, dont nous ne traiterons pas, l'expérimentateur devrait suivre les instructions particulières de la méthode à utiliser, consulter un statisticien et s'appuyer sur tous les principes généraux que nous exposons ici.

16. Lors d'essais ultérieurs avec des sols, on pourrait établir une série de dilutions avec du sol non contaminé, ce qui permettrait d'estimer des CI p ou des CE 50.

3.1 Effets quantiques

La mortalité est l'effet le plus commun auquel aboutissent les essais à concentration unique, et les données résultantes sont quantiques. Un essai pourrait servir à évaluer la mortalité d'amphipodes ou de larves de chironomes exposées à un sédiment non dilué ou la mortalité de truites arc-en-ciel dans un effluent non dilué. La partie supérieure gauche de fig. 3 montre le choix de tests statistiques en la circonstance.

3.1.1 Un seul échantillon sans répétition

Soumettre un échantillon à l'essai, sans répétition, est chose courante pour un rejet faisant l'objet d'une surveillance périodique. En employant un seul échantillon non réitéré et un témoin, on peut comparer le nombre d'organismes morts au moyen de l'un des tests exposés dans l'alinéa qui suit. Le test de comparaison devrait prendre la forme d'un test unilatéral de signification, parce que, normalement, l'expérimentateur ne se soucierait que d'une mortalité plus grande à la concentration d'essai que dans le milieu témoin ou de référence [17]. Parce que les tests se fondent sur des données limitées, on peut seulement s'attendre à ce qu'ils décèlent des effets relativement importants.

Voici les deux méthodes conseillées. L'annexe G fournit des exemples et renvoie à des publications, bien que ces tests sur des proportions soient traités dans les manuels classiques de statistique. Dans les comparaisons employant ces méthodes, l'hypothèse nulle est celle selon laquelle la concentration d'essai n'entraîne pas d'effet « pire » que les performances observées chez le témoin, c'est-à-dire un test unilatéral, mentionné précédemment. Les méthodes donnent de bons résultats, que le témoin manifeste ou non des performances réduites (par ex. une certaine mortalité).

- On recommande en premier lieu la ***méthode exacte de Fisher,*** parce que c'est, en effet, une méthode *exacte.* Elle demande peu de calculs, dans des sélections et manipulations des données exposées point par point dans un tableau simple de deux

17. On a recours à un test bilatéral comme le khi-deux si le sens de la différence n'a pas d'importance ou si l'on ne peut pas faire de supposition à son égard avant le début de l'essai. Dans les essais de toxicité létale, cela serait rarement le cas, car la mortalité est plus grande dans l'échantillon que chez le témoin.

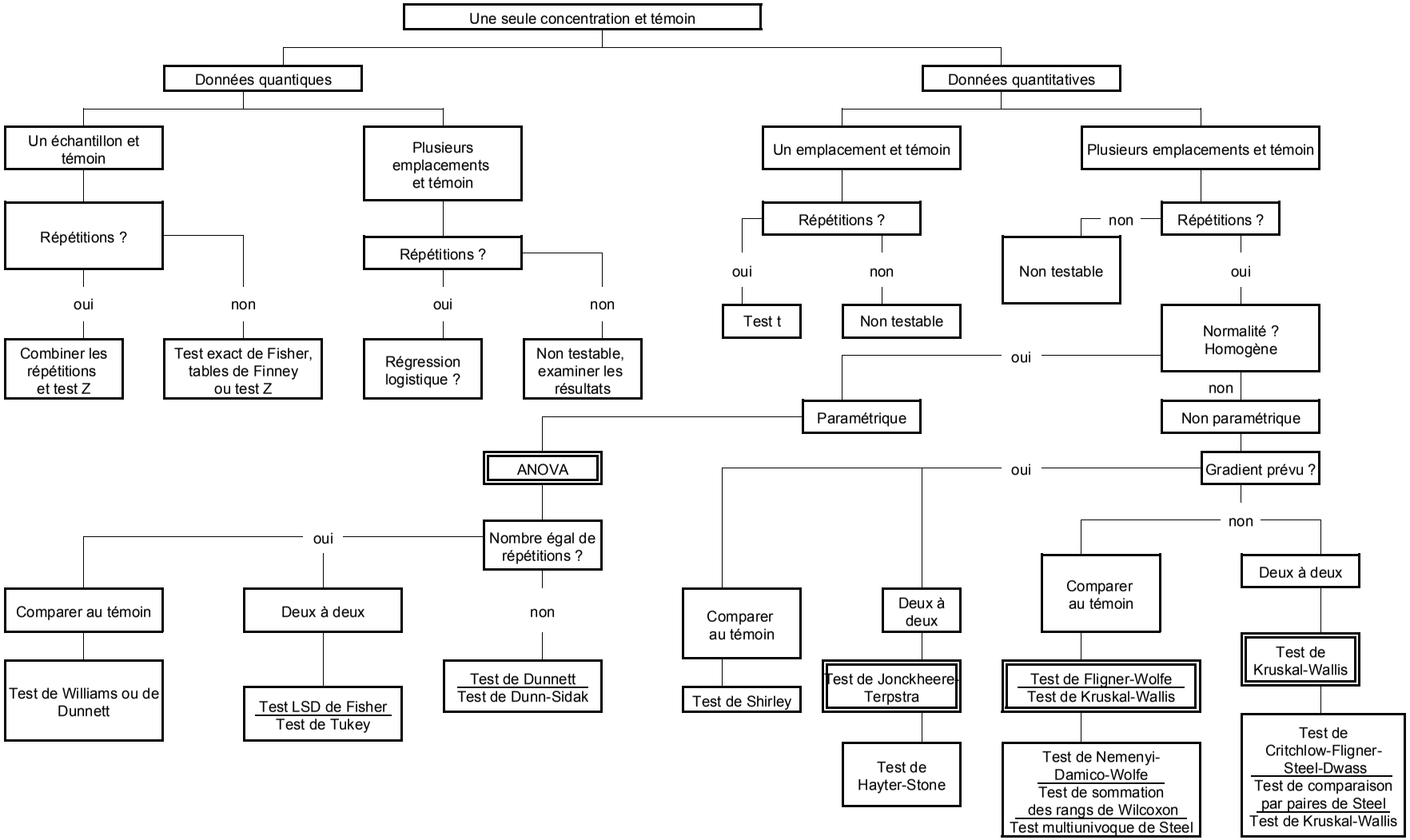


Figure 3. — Organigramme des méthodes statistiques applicables aux résultats de diverses catégories d'essais à concentration unique. Certaines de ces options pourraient être rarement utilisées. Les cases entourées d'un cadre double dénotent un test d'hypothèse nulle ; ce n'est que si cette hypothèse est rejetée que l'on passe à une comparaison multiple.

cases sur deux. On compare la valeur calculée à une valeur critique fournie, pour la méthode exacte de Fisher, dans les manuels de statistique générale.

- **« Tables de Finney ».** — Il s'agit simplement de comparer les données à des diagrammes, qui montrent immédiatement si l'effet expérimental est plus grand que celui que l'on observe chez le témoin. Les diagrammes sont présentés dans l'annexe G, mais ils ne valent que pour un nombre égal d'individus dans les enceintes expérimentales et les enceintes témoins, jusqu'à concurrence de 10. Dans les cas où le nombre d'individus n'est pas égal ou qu'il est supérieur à 10, on pourrait consulter la source de ces diagrammes, les tables publiées de Finney *et al.* (1963). Dans certaines bibliothèques universitaires, on pourrait trouver, mais difficilement, le recueil de ces tables. Les tables présentent les *p*-valeurs pour la comparaison de deux proportions, à la manière du test *t* [18].

Le ***test Z*** est une autre façon de comparer deux proportions. Nous ne le recommandons pas parce que les deux tests précités sont disponibles. Le test Z figure dans la plupart des manuels nord-américains de statistique [19] (par ex. Zar, 1999, p. 557), et nous donnons un exemple de son application dans l'annexe G. On compare une valeur calculée à une valeur critique de *Z*, qui, de fait, est trouvée dans des tables pour les valeurs de *t*. Le test se fonde sur l'approximation normale à la distribution binomiale, laquelle est indigente pour les petites tailles d'échantillon auxquelles s'appliqueraient les comparaisons dont nous parlons. L'approximation est particulièrement mauvaise quand les proportions observées se situent à l'extérieur de l'intervalle d'environ 0,4 à 0,6.

Ni la méthode de Fisher ni celle de Finney ne devrait faire l'objet d'une interprétation trop étroite. Pour la méthode exacte de Fisher, l'ordinateur produit habituellement la valeur de probabilité exacte. Même s'il fallait consulter les tables pour connaître les valeurs critiques de Z, l'expérimentateur serait néanmoins capable de juger de la *p*-valeur approximative. À vue de nez, la signification des *p*-valeurs dans l'intervalle général de 0,025 à 0,075 pourrait être considérée comme non concluante. Dans les études importantes, on pourrait effectuer des essais supplémentaires ou on devrait consulter un statisticien pour connaître les autres options possibles. S'il existait des tests statistiques de rechange, on devrait choisir ceux qui correspondent aux caractéristiques de l'essai particulier de toxicité que l'on réalise.

3.1.2 Répétition au même emplacement

Un programme d'essais à concentration unique pourrait parfois employer des échantillons réitérés du même emplacement, c'est-à-dire plusieurs échantillons prélevés en même temps et au même endroit. Cela serait plus probable dans les programmes d'étude de sols ou de sédiments que dans la surveillance des effluents liquides. Aucune méthode statistique usuelle n'est établie pour utiliser toute la gamme des données sur les effets quantiques, mais il subsiste des options. Dans cette situation, la méthode exacte de Fisher est toujours appropriée ; cependant, on devrait vérifier l'égalité des répétitions (avec la méthode exacte de Fisher) avant de grouper les données. Si le test montre que les données ne peuvent pas être groupées, l'expérimentateur doit s'interroger sérieusement sur la raison pour laquelle les effets sont significativement différents en un endroit. Une autre analyse possible combinerait les données des répétitions et soumettrait les proportions à un test Z, comme il est mentionné dans le § 3.1.1.

3.1.3 Lieux de prélèvement d'échantillons multiples

Si on soumettait des échantillons uniques (par ex. de sédiments) d'un certain nombre d'emplacements à un essai à une concentration avec témoin, il serait presque impossible de soumettre tout l'ensemble de données à un test statistique. D'habitude, une telle étude serait

18. On peut comparer deux proportions, comme, notionnellement on peut comparer deux moyennes. Quand on compare deux moyennes, on se sert d'une *distribution t* ou ou d'une *distribution normale* pour déterminer quelle différence entre deux moyennes est statistiquement significative. On peut déceler des différences toujours plus petites entre les moyennes à mesure que la taille de l'échantillon s'accroît et que s'amenuise la variabilité associée aux moyennes. On peut, par la même approche, comparer deux proportions, mais en employant la *distribution binomiale.* Les calculs décrits dans Zar (1999) sont quelque peu fastidieux. Il semble que Finney *et al.* (1963) aient comparé directement de la sorte deux proportions pour construire leurs tables et le diagramme de l'annexe G.

19. Dans les manuels européens, *Z* symbolise la variable normale centrée réduite, et on en trouve les valeurs dans les tables de la distribution normale.

préparatoire. On pourrait examiner les résultats pour y découvrir les signes d'un effet puissant et on pourrait poursuivre l'échantillonnage et les essais avec répétitions (voir le texte qui suit).

Des ensembles de données pourraient se prêter à des analyses spéciales, en consultation avec un statisticien. On pourrait appliquer l'analyse des observations aberrantes à l'identification de tout effet qui serait plus grave que celui qu'aurait subi le témoin et qui correspondrait aux échantillons de faible toxicité (§ 10.2). Si les emplacements constituaient un gradient (par ex. de l'amont vers l'aval), une régression permettrait de déceler l'effet de gradient.

Sous-échantillons de chaque échantillon. — On pourrait subdiviser les échantillons uniques de sédiment, de sol ou de liquide provenant chacun de plusieurs emplacements ainsi qu'un témoin ou un échantillon de référence en sous-échantillons et les soumettre à un essai. Cela représenterait une « répétition en laboratoire ». Pour ce qui concerne les effets quantiques, de telles données limitent les options qui s'offrent à l'analyse statistique (voir le texte qui suit). La répétition en laboratoire donne une idée de la variation dans les essais de toxicité effectués au laboratoire et de l'homogénéité de l'échantillon. Si la variation consécutive à la création de sous-échantillons était très faible, la répétition pourrait aider à distinguer les échantillons de terrain entre eux. Par exemple, si la variance des sous-échantillons était proche de zéro, ce serait le signe d'une bonne homogénéité des échantillons et d'essais de toxicité aux résultats précis ; on remarquerait les toxicités différentes des échantillons de terrain. Cependant, la variation de l'échantillonnage sur le terrain à un emplacement donné resterait inconnue, de sorte que les sous-échantillons ne procureraient pas la puissance nécessaire pour juger des différences entre les emplacements. C'est pourquoi on ne recommande pas particulièrement des répétitions en laboratoire à moins que cela ne soit spécifiquement voulu pour estimer la variation intralaboratoire. En général, il serait plus utile de consacrer l'effort supplémentaire aux échantillons réitérés (v. le § 2.5.2). Les conclusions des analyses statistiques portant sur les répétitions de laboratoire devraient être prudentes, et il faut les formuler sans ambiguïté. Sinon, on risquerait de mal interpréter les conclusions statistiques, en inférant, à tort, que toute différence décelée a résulté des différents emplacements sur le terrain.

Répétition sur le terrain. — Si on a prélevé des échantillons réitérés, c'est-à-dire plusieurs échantillons au même endroit, des analyses statistiques utiles deviennent faisables, même pour les données quantiques obtenues à une concentration. Une approche possible serait la régression logistique (§ 6.5), réalisée par un statisticien ou un toxicologue versé dans la statistique. La régression serait « catégorique », c'est-à-dire fondée sur le témoin, l'emplacement 1, l'emplacement 2, etc., plutôt que d'être la régression familière sur une variable indépendante continue comme la concentration. L'approche de la régression logistique pourrait être particulièrement fructueuse si on s'attend à un gradient d'effet (par ex. à différents endroits en « en aval » d'une source de pollution).

3.2 Effets quantitatifs à un endroit

Un exemple d'essai à concentration unique pour l'étude des effets quantitatifs serait la mesure du poids moyen des larves de chironomes après exposition à un échantillon de sédiment non dilué, par rapport au poids des larves exposées à un sédiment de référence ou à un sédiment témoin (EC, 1997a). On pourrait, en théorie, effectuer des essais préliminaires dans des enceintes expérimentales uniques, mais les essais définitifs porteraient sur des échantillons réitérés. La riche arborescence des choix est représentée dans les parties droite et inférieure de la fig. 3.

Sans répétition. — Si un seul échantillon était soumis à l'essai et s'il n'y avait qu'une seule matière témoin ou matière de référence, sans répétition, on ne pourrait comparer les résultats par aucun test statistique.

Répétition et comparaison par un test *t*. — Dans un essai quantitatif avec répétition auquel on soumettrait la matière à l'étude et la matière témoin ou de référence, un ***test t*** ordinaire conviendrait à l'analyse statistique. Ici encore, l'expérimentateur chercherait une taille réduite dans la matière à l'étude, de sorte que la valeur critique du test *t* serait celle d'un test unilatéral. La méthode suivie pour les tests *t* est fréquemment présentée dans les manuels de statistique ainsi que dans les logiciels tels que TOXSTAT.

Comme il en a déjà été question (§ 3.1.3), si les répétitions étaient des sous-échantillons d'un seul échantillon (« répétitions de laboratoire »), les conclusions du test statistique ne refléteraient que la variation intralaboratoire. On ne pourrait tirer aucune conclusion sur les différences du monde extérieur, par ex. si le lieu de prélèvement des échantillons diffère de celui d'où provient le témoin. Si, cependant, on avait utilisé des échantillons réitérés, les conclusions s'appliqueraient au monde réel, au moment et au lieu du prélèvement.

On peut appliquer le ***test* t** à la plupart des ensembles de données. Il fonctionne pour les nombres inégaux de répétitions dans l'essai et le témoin. À proprement parler, il repose sur l'hypothèse d'une distribution t de Student et de l'égalité des variances dans les deux groupes. Dans le doutes sur ces hypothèses, on pourrait vérifier la distribution t au moyen d'un graphique quantile-quantile ou, si la taille de l'échantillon était supérieure à 30, par un test de normalité. On pourrait tester l'homogénéité de la variance par les tests d'***O'Brien, de Levene*** ou ***de Bartlett*** ou par le ***test F*** (§ 7.3.1) [20]. Cependant, le test *t* est assez robuste, particulièrement si les échantillons et le témoin sont de tailles égales ou presque égales et si les nombres d'individus ne sont pas trop petits. Diverses modifications sont disponibles, et CETIS offre le *test* t *pour échantillons appariés*, le *test* t *pour variances égales* et le *test* t *pour variances inégales*.

3.3 Essais quantitatifs sur échantillons provenant de plusieurs endroits

Dans un autre type d'essai à concentration unique, on soumet les échantillons provenant de plusieurs endroits à un essai simultané, selon le même mode opératoire et avec la même matière témoin ou de référence. On procède généralement ainsi avec des échantillons de sols de divers endroits entourant un lieu contaminé ou de sédiments provenant de plusieurs emplacements dans un port, afin de délimiter une éventuelle zone de forte contamination. Il existe un guide approfondi sur le prélèvement et la répétition d'échantillons de sédiments (EC, 1994), que nous recommandons. Il existe également un guide de méthodes individuelles d'essai d'un sédiment telles que la méthode employant des vers polychètes (EC, 2001a).

Nous prenons comme exemple l'échantillonnage d'un sédiment sur le terrain en différents endroits. Les observations concernent les essais portant sur des effets quantitatifs tels que la modification du poids des organismes.

L'analyse statistique fructueuse d'échantillons de sédiments provenant de plusieurs endroits exige le prélèvement d'échantillons séparés à chaque emplacement (c'est-à-dire des *échantillons réitérés*). Dans le § 2.5, on traite de la manière de répéter les échantillons. Pour ce qui est du test d'hypothèse, une solution de rechange qui ne permet pas de distinguer les emplacements consiste à prélever un échantillon de chaque station, puis à le subdiviser plus tard en sous-échantillons (dits « répétitions de laboratoire »). L'essai ne livrerait que des renseignements limités sur la différence ou non d'un échantillon particulier par rapport à un autre échantillon particulier. Il *ne permettrait pas un test d'hypothèse sur l'absence de différence entre les emplacements (stations d'échantillonnage)* [v. le § 2.5].

Cas particulier des gradients. — Si on s'attend à l'existence d'un gradient d'effets décroissants sur une série de points d'échantillonnage de plus en plus éloignés d'une source de pollution, on peut se servir de la régression comme forme de test d'hypothèse. L'hypothèse nulle est qu'*aucun gradient n'existe.* L'hypothèse alternative est qu'un gradient d'effets existe en fonction de la distance de la source. Le choix et l'emploi d'une technique appropriée de régression exigent les conseils d'un statisticien. Pour cette analyse, les répétitions sont inutiles ; cependant, les échantillons réitérés permettent de tester le manque d'ajustement et, aussi, de rendre la régression statistiquement plus puissante. Un statisticien pourrait utiliser des sous-échantillons (« répétitions de laboratoire ») pour

20. Le test F est le dernier choix, mais, si on l'utilise, on en trouve la méthode dans tous les manuels de statistique, qui comportent habituellement des tables des valeurs critiques de F. S'il y a 4 répétitions et si chacune correspond à un poids moyen d'organismes survivants, la variance se calcule à partir des 4 moyennes, ce qui donne une variance pour la matière à l'étude et une autre pour le témoin. F est le rapport entre la variance supérieure et la variance inférieure. Le nombre de degrés de liberté est égal au nombre de répétitions dans chaque cas moins un. Si le test *t* était invalide en raison de l'inégalité des variances, on utiliserait une formule modifiée de ce test. Des exemples pratiques sont exposés dans une annexe d'USEPA (1995).

réduire la variance de l'erreur, mais l'effort d'échantillonnage devrait se concentrer sur les échantillons réitérés.

3.3.1 Tests paramétriques

Si on peut utilement ordonner les stations d'échantillonnage selon un gradient, le commentaire formulé dans le texte qui précède immédiatement s'applique, et des conseils supplémentaires sont donnés dans l'alinéa qui suit. Si, ayant prélevé des échantillons réitérés, on ne s'attend pas à l'existence d'un gradient, on pourrait effectuer un test d'hypothèse.

Pour le test d'hypothèse, le choix pour l'analyse statistique porte sur l'analyse de variance si les résultats répondent aux exigences de l'analyse paramétrique (§ 7.3). Si on compare chaque station d'échantillonnage à une matière de référence ou témoin, l'analyse de variance serait suivie du ***test de Dunnett*** (§ 7.5.1), ordre que nous préconisons. Certains vieux logiciels pourraient exiger un nombre égal de répétitions pour le test de Dunnett, mais les plus récents se sont libérés de cette contrainte (v. le § P.4.2 de l'annexe P). Le ***test de Williams*** pourrait remplacer le test de Dunnett, s'il existait visiblement un gradient d'effets tel qu'en une série de points se succédant vers l'aval à partir d'une source de pollution et si on se servait d'un test d'hypothèse. Le test de Williams comparerait les effets à chaque point à ceux de la station témoin, mais il tiendrait compte de la nature ordonnée des points, ce qui procurerait une analyse plus sensible (v. le § 7.5.1).

En principe, l'expérimentateur pourrait vouloir savoir quels points d'échantillonnage diffèrent des autres. Dans ce cas, il pourrait soumettre à l'essai, sans en diluer aucun, plusieurs échantillons de terrain du point ou de chacun des points en amont du point de rejet de l'effluent et, de même, d'un certain nombre de points en aval. Il pourrait vouloir effectuer des comparaisons deux à deux, dans le cadre d'une étude élargie des points en question, comme s'il est possible ou non de distinguer de la station en amont un point en aval montrant le « meilleur » retour à la normale. Pour effectuer une telle évaluation, une analyse de variance pourrait être suivie de la ***méthode LSD*** (*Least Significant Difference*, plus petite différence significative) ***de Fisher*** ou du ***test de Tukey.***

La méthode LSD est utile à la comparaison deux à deux à l'intérieur d'un ensemble plus grand de données parce qu'elle est relativement facile à réaliser et que l'on peut l'étendre aux cas comptant un nombre inégal de répétitions. Elle ne fait généralement pas partie des progiciels utilisés pour l'étude de la toxicité, mais, dans le § 7.5.1, on donne des conseils sur son emploi. D'autres conseils sur les tests paramétriques de comparaisons multiples sont donnés dans le § P.4 de l'annexe P.

Répétitions inégales. — Comme nous l'avons déjà mentionné, le test de Dunnett accepte les nombres inégaux de répétitions, dans les progiciels modernes de statistique qui sont les plus susceptibles d'être trouvés dans les laboratoires. Les vieux progiciels de toxicologie pourraient n'avoir que la version capable de traiter les nombres égaux de répétitions. Si on ne dispose pas d'un programme capable de traiter les nombres inégaux, on pourrait lui appliquer une modification, qui est expliquée dans Newman (1995) et dont on présente des exemples pratiques dans USEPA (1995). Les autres options pour les nombres inégaux d'observations sont le ***test de Dunn-Sidak*** ou le ***test t ajusté de Bonferroni*** (v. le § P.4 de l'annexe P).

3.3.2 Tests non paramétriques

Si les résultats d'essais de toxicité sur échantillons multiples n'ont pas répondu aux conditions de normalité et l'homogénéité de la variance, on devrait utiliser des tests non paramétriques. Les options qui s'offrent alors occupent la partie inférieure droite de la fig. 3. Des commentaires pertinents et des détails plus nombreux sont offerts dans tous les alinéas du § 7.5, y compris des observations sur la disponibilité des tests.

L'une des ramifications des tests proposés s'impose si on s'attend à l'existence d'un ordre ou gradient dans les résultats et si chaque emplacement sera comparé au témoin. Le ***test de Shirley*** pourrait être utilisé pour faire ces comparaisons (Shirley, 1977). Si on s'attend à l'existence d'un ordre et si on veut effectuer une comparaison deux à deux (chaque emplacement avec chacun des autres), on pourrait utiliser le ***test de Jonckheere-Terpstra*** pour tester l'hypothèse nulle de l'absence de différence (Jonckheere, 1954). Si l'hypothèse est rejetée, on passe au test de comparaisons multiples ***de Hayter et Stone*** (1991).

Une autre ramification des tests proposés entrerait en jeu si on ne s'attend à aucun ordre de concentration ou d'effet dans l'ensemble des résultats d'essais sur échantillons multiples. On pourrait alors comparer les effets à ceux qu'a éprouvés le témoin en testant l'hypothèse nulle de l'absence d'effet, au moyen du test de ***Fligner et Wolfe*** (1982). Si ce test n'était pas accessible, celui de ***Kruskal et Wallis*** (1952) ferait l'affaire. Si on rejette l'hypothèse nulle, on pourrait passer à un test de comparaisons multiples. Le premier choix irait au test de ***Nemenyi-Damico-Wolfe*** (Damico et Wolfe, 1987). Sinon, le deuxième choix serait le **test de la somme des rangs de Wilcoxon** ou le ***test multiunivoque de Steel*** (Steel, 1959).

On pourrait également vouloir effectuer une comparaison deux à deux (chaque emplacement l'un avec l'autre) si on ne s'attend à aucun ordre dans les effets. On testerait d'abord l'hypothèse nulle (absence d'effet de l'emplacement) par le test de Kruskal-Wallis. Si on concluait à l'existence d'une différence, on se servirait du ***test de Critchlow-Fligner-Steel-Dwass*** (Critchlow et Fligner, 1991) pour déterminer la ou les différences. Dans le cas contraire, on pourrait utiliser le ***test de comparaison par paires de Steel*** (Steel, 1960) pour données équilibrées (nombre égal de répétitions) ou, de nouveau, le test de Kruskal-Wallis, cette fois comme test de comparaisons multiples pour données non équilibrées.

Essais quantiques pour estimer la CE p

À la fin d'un essai de toxicité quantique, chaque organisme présente ou ne présente pas l'effet défini. L'effet est binaire : un ver de terre meurt ou vit, un œuf est fécondé ou reste non fécondé, un poisson manifeste ou non une réaction d'évitement. Dans ce contexte, *binaire* et *quantique* sont synonymes. Ainsi, la plupart des essais quantiques (de toxicité) se fondent sur la proportion d'organismes manifestant l'effet après l'exposition à une concentration fixe et d'une durée définie de la matière à l'étude.

Les résultats quantiques suivent une *distribution binomiale,* qui détermine le choix des tests statistiques appropriés. L'expérimentateur en quête de plus de renseignements dans un manuel de statistique devrait consulter les parties ou chapitres sur les données binaires et la distribution binomiale. Collett (1991) décrit les méthodes d'analyse des données binaires et il fait remarquer que les techniques bien connues de l'analyse de variance et de la régression linéaire simple sous les formes utilisées avec des données continues (quantitatives) ne sont pas utilisables directement avec des données quantiques (v. la fin du § 4.3). Des méthodes bien établies permettent d'ajuster les modèles aux données quantiques, mais les méthodes permettant de vérifier l'ajustement sont moins bien établies. (Tout en assimilant les bons conseils de Collett en matière de statistique, le lecteur devrait se méfier des déclarations sur les questions toxicologiques qui pourraient sembler trompeuses à première vue, comme il est expliqué dans le § 2.3.1.) Sont également utiles les manuels de statistique de Finney (1971 ; 1978) et d'Ashton (1972), lequel se concentre sur la modélisation logistique linéaire, particulièrement appropriée aux résultats des essais de toxicité quantiques. Hosmer et Lemeshow (2000) sont des auteurs plus récents à avoir traité de la régression logistique. Fleiss (1981) aborde certains aspects de la question tels que les tableaux de contingence.

La plupart des essais quantiques d'écotoxicité se fondent sur l'effet létal aigu. Bien qu'ils ne permettent pas d'estimer une concentration « inoffensive », ils ont joué un rôle de longue date dans l'évolution de l'écotoxicologie et ils ont permis de réunir une masse importante de résultats. Les essais quantiques (de toxicité) continuent d'être très utilisés, particulièrement à des fins réglementaires, peut-être parce qu'ils utilisent souvent des espèces bien connues comme la truite arc-en-ciel. Ces essais possèdent des qualités telles que la rapidité, une économie raisonnable, l'estimation d'un paramètre de toxicité non ambigu, un effet manifestement nuisible. Ils permettent de comparer la toxicité relative de matières ou la sensibilité d'espèces et ils permettent l'examen initial de la toxicité ou la surveillance des modifications survenant dans un effluent [21]. Conjointement à ces essais, il existe des méthodes bien établies d'analyse statistique. Certains essais de toxicité sublétale sont également quantiques et utilisent les mêmes techniques d'analyse.

L'organigramme général de l'analyse est raisonnablement simple (fig. 4). On recommande pour l'utilisation en routine la régression logit ou probit (souvent appelée « analyse par la méthode des probits ») si les données s'y prêtent, parce qu'un long historique d'emploi signifie que des programmes éprouvés et commodes d'analyse sont facilement accessibles. Si les données ne satisfont pas aux exigences de cette régression, des méthodes de rechange, bien que théoriquement moins attrayantes, sont capables de traiter les données auxquelles on a couramment affaire (§ 4.3).

Nous formulons, dans la présente section, des observations sur les étapes de l'analyse statistique des

21. Les essais de toxicité létale ne sont pas nécessairement inférieurs aux essais de toxicité sublétale ; parfois, ils sont précisément l'outil dont on a besoin pour l'étude. Un exemple d'emploi de la toxicité létale pour examiner des sujets scientifiques complexes est l'outil puissant que constituent les *relations quantitatives structure-activité* (QSAR ou RQSA), c'est-à-dire les relations entre la structure chimique des substances et leur toxicité pour les organismes aquatiques. Grâce à des programmes de recherche d'envergure, on a défini, au moyen d'essais de toxicité létale, une série de QSAR complexes permettant de formuler des prévisions efficaces sur les nouvelles substances dangereuses aux structures chimiques semblables (Broderius, 1991 ; USEPA, 1994e).

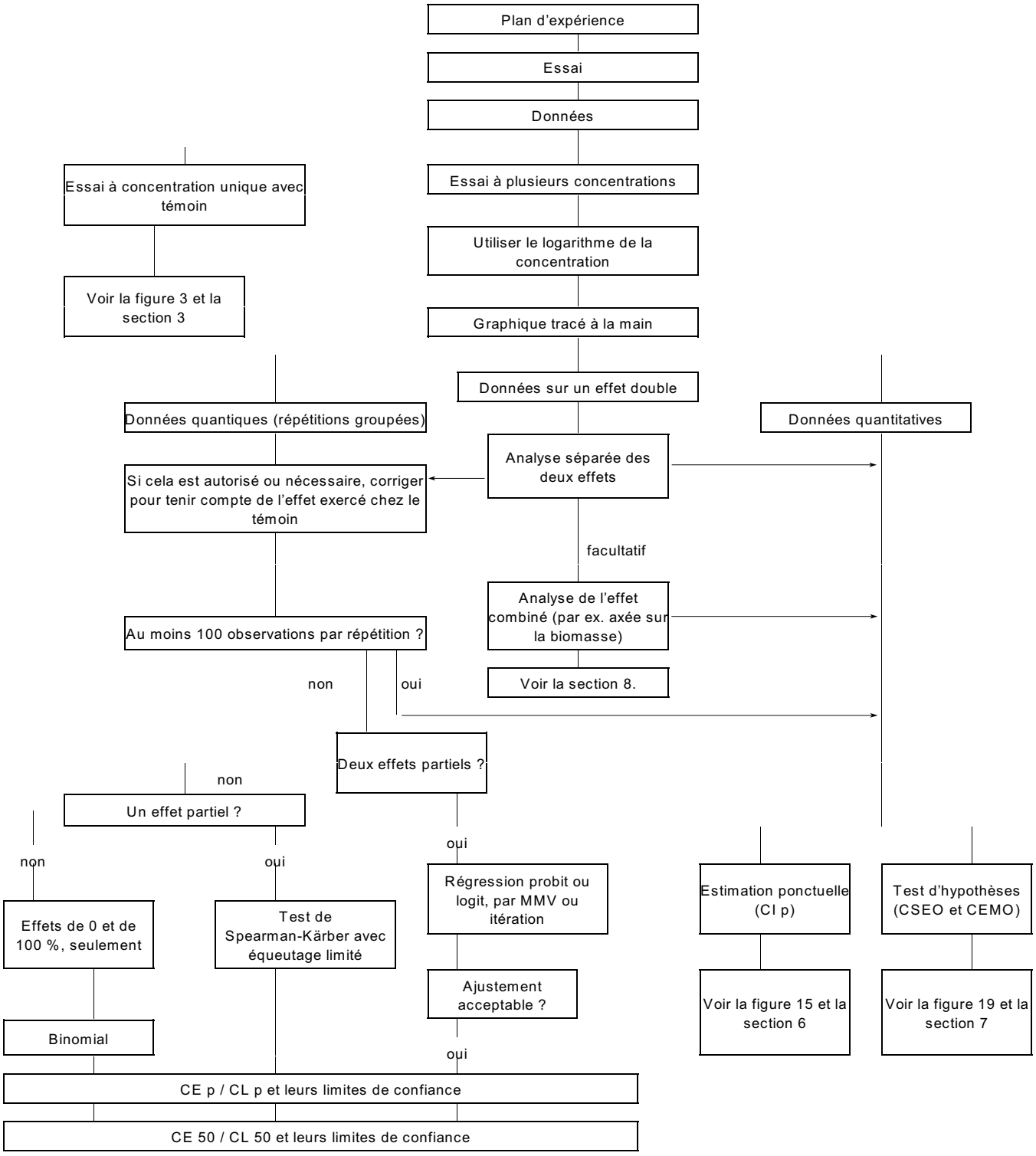

Plan d'expérience
Essai
Données
Essai à concentration unique avec témoin
Essai à plusieurs concentrations
Utiliser le logarithme de la concentration
Voir la figure 3 et la section 3
Graphique tracé à la main
Données sur un effet double
Données quantiques (répétitions groupées)
Données quantitatives
Analyse séparée des deux effets
Si cela est autorisé ou nécessaire, corriger pour tenir compte de l'effet exercé chez le témoin
facultatif
Analyse de l'effet combiné (par ex. axée sur la biomasse)
Au moins 100 observations par répétition ?
Voir la section 8.
non
oui
Deux effets partiels ?
non
Un effet partiel ?
oui
non
oui
Régression probit ou logit, par MMV ou itération
Estimation ponctuelle (CI p)
Test d'hypothèses (CSEO et CEMO)
Effets de 0 et de 100 %, seulement
Test de Spearman-Kärber avec équeutage limité
Ajustement acceptable ?
Voir la figure 15 et la section 6
Voir la figure 19 et la section 7
Binomial
oui
CE p / CL p et leurs limites de confiance
CE 50 / CL 50 et leurs limites de confiance

Figure 4. — Organigramme des méthodes d'analyse s'appliquant aux résultats des essais quantiques. Les méthodes quantiques occupent les branches de la partie inférieure gauche de l'organigramme.

données quantiques. Les méthodes diffèrent tout à fait de celles qui s'appliquent aux données quantitatives, discutées dans les sections 6 et 7. Certains essais engendrant beaucoup d'observations quantiques, sont analysables par des méthodes quantitatives (§ 6.1.1).

4.1 Les paramètres de toxicité estimés au moyen d'essais quantiques

Repères

- *Dans un essai de toxicité quantique, chaque organisme présente ou ne présente pas un effet. L'effet pourrait être létal ou sublétal (par ex. l'immobilisation).*
- *Dans un essai de toxicité quantique à plusieurs concentrations, ce paramètre est la* concentration efficace, *habituellement la* concentration efficace médiane ou 50 (CE 50). *Les essais de toxicité sublétale sont une sous-catégorie de ce type d'essai, et le paramètre habituellement estimé est la* concentration létale médiane ou 50 (CL 50). *La durée d'exposition doit être précisée, par ex. CE 50 après 96 h.*

Dans les essais quantiques, on expose, pendant une durée fixe, des groupes séparés d'organismes à une concentration faisant partie d'une série de concentrations fixes. Il est souhaitable de disposer d'un nombre égal d'organismes à chaque concentration, et la durée d'exposition doit être la même. Les observations sont le nombre d'organismes touchés à chaque concentration (par ex. le nombre d'organismes morts). Les proportions d'organismes touchés permettent les analyses statistiques appropriées. (Fleiss [1981] expose les rudiments des analyses pratiques des proportions).

Le paramètre de toxicité estimé au moyen d'un essai quantique est la *concentration efficace à laquelle se manifeste un effet toxique sur un pourcentage précisé d'organismes exposés, la concentration efficace à p %* (CE p). Le pourcentage choisi (*p*) est d'habitude 50 %, c'est-à-dire la *concentration efficace médiane, 50 % ou 50,* susceptible de causer un effet chez la moitié des organismes. Dans le langage courant, c'est une estimation de la concentration qui n'affecterait que l'organisme « typique » ou « moyen », un paramètre de toxicité possédant une certaine validité. Un motif supplémentaire pour choisir l'effet à 50 % est que lui correspond l'intervalle de confiance à son plus étroit, qui s'élargit à mesure que l'on s'éloigne de la médiane, de sorte qu'il serait très large si on choisissait un pourcentage d'effet extrêmement petit ou grand (§ 4.2.4). Les estimations de la CE 25 ou de la CE 20 sont « en demande », et l'on peut également estimer ces paramètres à l'aide de certains programmes d'analyse (§ 4.2.5).

Il faut toujours préciser la durée d'exposition avec la CE p, par ex. comme suit : « la CE 50 après 96 heures (souvent notée CE 50 96 h) ». Les essais quantiques sont généralement associés à des expositions de courte durée. La CE 50 de la viabilité des œufs de salmonidés après une exposition de 7 jours, par ex., résulte d'un essai de toxicité aiguë parce que ce dernier se déroule sur une petite fraction du cycle vital de l'organisme. Moins généralement, un essai quantique pourrait être de toxicité chronique, par ex. un essai de la mortalité chez le poisson après des mois d'exposition.

La CE p s'applique à tout effet quantique, létal ou sublétal. Elle englobe une sous-catégorie fréquemment utilisée, la concentration létale (CL p), presque toujours la CL 50). Dans le texte qui suit, les notions de CE p ou de CE 50 engloberont de façon plus générale la notion de CL 50 [22, 23].

Pour chaque CE p, il faudrait préciser les limites de confiance (v. le § 4.2.4).

22. Parfois, il est difficile de déterminer la mort chez un animal, particulièrement un invertébré. Un paramètre convenable de toxicité peut être la CE 50 pour l'immobilisation, comme dans l'essai d'Environnement Canada employant des daphnies (EC, 1990b). Cette concentration est écologiquement significative et devrait être acceptée ; on pourrait l'utiliser chez d'autres types d'organismes.

23. *CL 50, CE 50, CI p, CI 25* etc. sont tous des noms ou substantifs. Il est inutile d'écrire « valeur de la CL 50 » ou « estimation de la CI p », expression pléonastiques. La structure de la phrase devrait s'adapter à ce à quoi ces abréviations tiennent lieu. Dans l'abréviation *CE 50,* le mot sous-entendu est le substantif *concentration.* On n'écrirait pas « valeur de la concentration » ni, maladroitement, « valeur de la CE 50 ». Une maladresse encore plus évidente, que l'on lit de temps à autre, est « concentration de la CL 50 », ce qui revient à écrire « concentration de la concentration létale médiane ».

4.2 Marche à suivre pour toutes les méthodes d'estimation d'une CE p

Repères

- *Il est impossible d'estimer, par quelque méthode que ce soit, la CE 50 si, à au moins une concentration, il n'existe pas un effet d'au moins 50 %. On peut estimer la CE 50 s'il existe un effet nul à une concentration, mais que, à toutes les concentrations supérieures les effets sont d'au moins 50 %, et si un effet linéaire logique est évident. L'estimation de la CE 50 est plus fiable si des effets partiels encadrent cette concentration. Cependant, on peut interpoler la CE 50 à partir des effets à 0 et à 100 % manifestés à des concentrations successives ; cette estimation pourrait être précise si ces concentrations sont rapprochées.*
- *Dans toute estimation, les renseignements tirés des concentrations ne provoquant aucun effet ou provoquant un effet complet devraient servir à établir la position et la pente de la relation dose-effet, mais on ne peut utiliser à cette fin qu'un seul effet nul et qu'un seul effet à 100 %.*
- *Il faut porter les concentrations sur une échelle logarithmique, conformément à l'hypothèse scientifique formulée dans le choix des concentrations. Cette échelle permet habituellement de supprimer l'asymétrie, pour faciliter l'appréciation visuelle de l'ajustement. La représentation du pourcentage d'effet sur une échelle logit ou probit complète habituellement le redressement de la courbe sigmoïde en une droite.*
- *Pour estimer les paramètres de toxicité au moyen de programmes informatiques, il faut des contrôles pour vérifier la saisie fidèle des observations et la plausibilité des résultats livrés par le programme. L'un de ces moyens de contrôle serait un graphique tracé à la main du pourcentage d'effet en fonction du logarithme de la concentration, dont nous donnons des exemples. On devrait comparer le graphique et l'estimation de la CE p qu'il permet à ceux que produit l'ordinateur.*
- *Les estimations des paramètres tels que la CE 50 sont normalement calculées sous forme de logarithmes, puis, dans un souci d'intelligibilité, converties en valeurs arithmétiques de la concentration. Avant de calculer la moyenne ou d'effectuer toute autre opération mathématique sur les CE 50, il faut convertir de nouveau ces dernières en logarithmes. Le temps, également, doit être exprimé selon une échelle logarithmique.*

Certaines règles générales, incontournables, s'appliquent à toutes les méthodes d'estimation de la CE p. Les programmes informatiques ne prémunissent pas nécessairement contre les inobservations qui pourraient provoquer des erreurs d'analyse.

- On combine les résultats obtenus pour des répétitions (récipients) correspondant à une concentration donnée[24].

- Si à au moins une concentration ne correspond pas un effet d'au moins 50 %, on ne peut pas estimer la CE 50. (Bien sûr, on peut affirmer que la CE 50 est supérieure à la concentration maximale utilisée dans l'essai.) On ne peut pas, par extrapolation, estimer une concentration qui causerait un effet de 50 % à partir d'une concentration exerçant un effet de moins de 50 %. Il se peut que des effets de 50 % ou plus ne se manifestent jamais à des concentrations plus fortes : par ex. un toxique pourrait atteindre sa limite de solubilité et ne pas exercer davantage sa toxicité ou, encore, les organismes survivants pourraient *tolérer* de fortes concentrations.

L'expérimentateur doit pouvoir compter sur des méthodes de repli pour l'analyse, parce que beaucoup d'ensembles de résultats n'englobent pas les deux effets partiels dont on a besoin pour la régression logit ou probit. Le personnel des laboratoires d'Environnement Canada a estimé que jusqu'à 90 % des essais réglementaires et des essais de surveillance usuels ont

24. Si les résultats étaient portés sur un graphique tracé à la main pour obtenir une estimation graphique de la CE 50, on pourrait distinguer les répétitions pour se donner une impression visuelle de leur variation. Les méthodes informatisées usuelles d'estimation de la CE p combinent les répétitions. Dans l'avenir, des systèmes mathématiques d'analyse pourraient utiliser les répétitions séparées, mais, actuellement, peu de progiciels sont capables d'utiliser correctement cette information.

pour résultats une mortalité partielle ou n'en ont aucune (Doe, 1994) et que, en conséquence, ces résultats ne peuvent pas être traités par une régression probit ou logit. De même, on trouve dans APHA *et al.* (1992) un exemple selon lequel, sur 60 essais de toxicité aiguë en milieu aquatique, seulement 4 (7 %) ont donné des résultats qui satisfaisaient aux hypothèses et aux exigences de la régression probit en matière de données. Il est souvent très important, pour les besoins de la surveillance, que les estimations de la CE 50 et leurs limites de confiance soient acceptables, même si, du point de vue statistique, elles ne sont pas parfaites. Les méthodes de repli permettent habituellement d'obtenir l'estimation acceptable. Souvent, la répétition de l'essai pour obtenir un résultat plus précis ou mieux défendable est impraticable, parce que l'échantillon est soit épuisé, soit trop vieux.

L'estimation de la CE p peut habituellement être considérée comme plus fiable si les données révèlent un effet partiel sous la CE p et un autre au-dessus. Néanmoins, on peut interpoler la CE 50 sans connaître d'effets partiels, si une concentration provoque un effet nul et que la concentration supérieure suivante provoque un effet total (100 %) [V. le § 4.5.7, sur la méthode binomiale). En effet, un tel essai à seulement deux effets (nul et total) permettrait une approximation excellente de la CE 50 si les concentrations étaient raisonnablement rapprochées. Les lignes directrices qui suivent concernent les effets nul et total.

- Il est licite d'estimer une CE p (par ex. la CE 25) à partir de données comprenant un effet nul, mais aucun effet partiel au pourcentage *p* choisi ni à un pourcentage inférieur. Il faut observer, au-dessus de *p* %, une croissance régulière des effets, compatible avec une relation linéaire, et la droite ajustée doit décrire une proportion statistiquement significative de la variabilité totale [25]. Certaines autorités en la matière et des programmes informatiques pourraient appliquer des exigences plus rigoureuses à l'égard de l'estimation d'une CE p [26]. À notre avis, cependant, il n'est pas nécessaire de ne pas tenir compte des résultats d'un essai à cause d'une absence d'effet partiel sous le pourcentage d'effet choisi ou à ce pourcentage.

- Si une concentration ne correspond à aucun effet, on devrait utiliser ce fait dans l'ajustement de la droite. De même, on devrait utiliser un effet de 100 %. Ces observations sont de peu de poids dans l'ajustement d'une droite effet-concentration, mais elles aident à établir la pente.

- Si, toutefois, des concentrations successives donnent une série d'effets nuls ou une série d'effets à 100 %, on ne devrait utiliser que les concentrations les plus centrales de la série pour l'estimation de la CE 50 (Ashton, 1972). Autrement dit, la concentration utilisée devrait être la concentration maximale des concentrations successives ayant donné un effet nul ou la concentration minimale ayant donné un effet de 100 %. Dans chaque cas, la concentration (et l'effet) à utiliser est celle qui est la plus « rapprochée du centre » de la distribution de données. L'emploi d'un seul effet nul et (ou) d'un effet à 100 % est important pour les analyses informatisées. Si l'expérimentateur saisit plus d'une valeur successive à 0 ou à 100 %, le programme s'efforce d'utiliser la ou les valeurs supplémentaires, il modifie la pente et la position de la droite ajustée et, de la sorte, aboutit à des estimations quelque peu déviantes de la CE 50 et de ses limites de confiance. La solution à ce problème ne consiste pas à saisir des valeurs « supplémentaires » dans le programme. Ce point est important, et

25. Cette croissance doit être prouvée graphiquement. La valeur de khi-deux ne devrait pas excéder la valeur critique lorsqu'une droite est ajustée par la régression probit, condition qui s'applique à tous les tests (v. le § 4.5.4).

26. Le programme informatique de Stephan *et al.* (1978) pour la régression probit (§ 4.5.3) exige deux effets partiels, à l'instar de tous les programmes de transformation probit. En outre, il exige : soit *a*) au moins un effet sous 50 % et au moins un au-dessus ; soit *b*) un effet à 50 % et au moins un autre, soit au-dessous, soit au-dessus de 50 %. Ces exigences sont raisonnables, bien que légèrement plus rigoureuses que les recommandations actuelles d'Environnement Canada.

Certains programmes informatiques pourraient estimer une CE 50 à partir de données insuffisantes, mais on ne devrait pas accepter le résultat, à moins de satisfaire aux exigences d'une méthode d'essai d'Environnement Canada. Par exemple, le programme informatique de Hubert (1987) permet d'obtenir des estimations à partir de deux effets inférieurs (ou supérieurs) à 50 %. Dans le premier cas, cela n'est pas acceptable pour Environnement Canada, parce qu'aucune donnée ne prouve que les effets atteindraient le taux de 50 %. Dans le second cas (deux effets supérieurs à 50 %), cela ne serait acceptable pour Environnement Canada que s'il y avait un effet nul à quelque autre concentration inférieure.

l'erreur est fréquente. (Bien sûr, nos commentaires sur les pourcentages nuls successifs ne s'appliquent pas à un témoin.)

4.2.2 Transformation log-probit

En choisissant les concentrations d'exposition pour un essai, l'expérimentateur est presque certain de suivre la coutume en les choisissant dans une suite géométrique ou logarithmique. Cela est une admission tacite de l'adoption du logarithme de la concentration comme l'unité la plus pertinente pour exprimer la dose (le *dose metameter*), et c'est la base qui convient le mieux à l'analyse statistique ultérieure, comme nous l'expliquons dans le § 2.3. Une fois calculé, le paramètre de toxicité devrait être considéré comme un logarithme. Cependant, on convertit habituellement un paramètre de toxicité tel que la CE 50 en valeur arithmétique, pour aider à la compréhension quotidienne des chiffres. Un avantage concret du logarithme de la concentration est que ce dernier supprime d'habitude l'asymétrie des données représentées graphiquement (fig. H.1 de l'annexe H).

De même, le temps biologique est le mieux pris en compte en tant que phénomène logarithmique (§ 2.3.6). D'où l'emploi des logarithmes de temps et de concentration dans la construction des courbes de toxicité (§ 5.2) et la nécessité d'employer le logarithme du temps dans des calculs, si on doit estimer les temps efficaces 50 (TE 50) [§ 5.1].

L'emploi des probits pour représenter le pourcentage d'effet résulte de la recherche d'une façon de représenter les données au moyen d'une relation linéaire entre les données. Empiriquement, les probits rectifient une distribution sigmoïde des données sur les effets, ce qui était commode à l'époque précédant l'avènement des ordinateurs, parce qu'une droite était plus facile à analyser. L'habitude s'est maintenue dans les programmes informatiques modernes. Les probits « étirent » graduellement l'échelle verticale des effets qui s'éloignent du taux de 50 % (pour une représentation graphique, v. la fig. 5, et, pour des explications, l'annexe H).

Le graphique log-probit combiné représente sous la forme d'une droite ce qui, en réalité, est une courbe log-normale cumulée (annexe H ; Buikema *et al.*, 1982 ; chap. 1 de Rand et Petrocelli, 1985).

Pour le traçage à la main, il est commode de se procurer du papier graphique « log-probit » et de simplement y porter les valeurs arithmétiques. Si on ne peut pas se procurer de papier log-probit dans une papeterie, on peut photocopier le graphique vierge de l'annexe I.

Dans un souci de simplicité, les descriptions et les exemples présentés dans l'alinéa suivant concernent les probits, mais on pourrait utiliser les logits, et les mêmes observations générales s'appliqueraient. La seule exception serait que l'on peut acheter du papier log-probit, mais non du papier log-logit.

4.2.3 Estimation de la CE 50 à l'aide d'un graphique tracé à la main

Dans un premier temps, pour se faire une idée générale des données et de la CE 50 résultante, on devrait tracer à la main un graphique des résultats. Un groupe de travail constitué de statisticiens et de toxicologues (OECD, 2004) est d'accord. Il décrit « une analyse typique des données » et en énumère les étapes : « on commence par obtenir un tracé des données, que l'on apprécie visuellement ». En conséquence, nous décrivons d'abord les estimations graphiques. Ces estimations illustrent commodément quelques notions et difficultés. Dans une certaine mesure, un graphique explique ce qu'un programme d'analyse informatique effectue. La fig. 5 montre des exemples à l'aide d'ensembles représentatifs de données.

La plupart des méthodes d'Environnement Canada relatives aux essais quantiques et un guide de l'USEPA (2000a) recommandent une ligne ajustée à vue d'œil pour estimer une CE 50 approximative, vérifier le caractère « raisonnable » d'une estimation informatique. L'expérimentateur consciencieux devrait toujours tracer le graphique à la main afin d'appliquer le critère le plus utile d'évaluation de la validité : le bon sens. L'exercice pourrait révéler un ensemble irrégulier d'effets qui ne devraient pas être imposés à une analyse mathématique usuelle. Le tracé est particulièrement nécessaire si on a utilisé la méthode d'analyse statistique de Spearman-Kärber (§ 4.4 et 4.5.6). Le graphique tracé à la main peut servir de confirmation ou d'avertissement, mais il *ne donne pas* un paramètre de toxicité ayant une valeur définitive et déclarable.

Certains s'insurgent contre la nécessité de tracer un graphique à la main, à notre époque où l'infographie fait des merveilles, mais le crayon et le papier conservent leur importance. Une erreur dans la saisie des données dans l'ordinateur se reproduirait dans le graphique dessiné par l'ordinateur ainsi que dans l'estimation mathématique de la CE 50, et l'accord des deux moyens ne permettraient pas de déceler l'erreur de saisie [27]. Les expérimentateurs devraient, en effet, utiliser les programmes informatiques les plus modernes et les plus puissants auxquels ils ont accès (par ex. l'estimation du maximum de vraisemblance). Mais, d'un point de vue pratique, une vérification rapide au moyen d'un graphique tracé à la main pourrait se révéler la meilleure façon de remédier aux résultats erronés qui ont été évidents dans les programmes antérieurs d'essais d'Environnement Canada. Des erreurs peuvent être plutôt fondamentales et simples. Les nouveaux employés pourraient avoir besoin de temps pour acquérir des compétences dans les essais de toxicité et l'analyse statistique. Des gestionnaires de données pourraient saisir les résultats des essais sans bien comprendre le programme d'analyse ou sans savoir si son résultat est raisonnable. Les graphiques tracés à la main aident à remédier à ces situations.

On devrait comparer le graphique tracé par l'ordinateur au graphique tracé à la main. On pourrait aussi superposer les résultats de l'analyse statistique sur les données brutes ou le long de la droite ajustée à vue d'œil, comme moyen de vérification visuelle. Tout écart appréciable devrait faire l'objet d'une enquête et devrait être résolu. Dans la présente section, nous donnons des exemples de graphiques avec des conseils sur la façon d'ajuster les droites à vue d'œil. Avec la pratique, ces droites donneront des estimations de la CE 50 qui s'écartent de quelques pourcentages à peine des estimations de l'ordinateur, servant ainsi à remettre, comme on le désire, les pendules à l'heure.

Comme nous l'avons mentionné, les concentrations sont portées sur le graphique à une échelle logarithmique, et le pourcentage d'effet est sur l'échelle probit (fig. 5). Parce que l'échelle probit n'atteint jamais ni 0 ni 100 %, les valeurs extrêmes sont désignées par une flèche, comme dans les fig. 5A et 5B. La flèche pointe dans la direction où les valeurs réelles se trouveraient, au-delà des valeurs de 2 et de 98 % qui ont été fixées comme limites arbitraires sur ce papier quadrillé. En dépit de leur peu de poids, on devrait porter sur le graphique les effets à 0 et à 100 %, si on les connaît, parce qu'ils aident parfois à caler une ligne établie à partir d'un petit nombre de données. Ici encore, on ne devrait porter sur le graphique qu'une donnée pour une série d'effets nuls ou d'effets à 100 % successifs : la plus rapprochée du centre de la distribution.

En ajustant la droite des probits à vue d'œil, on devrait se servir d'une règle transparente, en la déplaçant ou en la faisnt tourner afin de réduire au minimum les distances *verticales* entre les points observés et la droite ajustée. En même temps, on devrait pondérer mentalement les points. On devrait affecter du plus grand coefficient de pondération les points les plus rapprochés de l'effet de 50 % et du plus petit coefficient les points correspondant à l'effet de 0 et de 100 %. À vue de nez, la pondération maximale devrait aller aux points situés entre les effets de 16 et de 84 %, qui se trouvent à ± 1 probit de la médiane. Une valeur de 10 ou de 90 % a un poids environ deux fois moins grand qu'une valeur située dans l'intervalle de 40 à 60 %. À 3 ou à 97 % d'effet, le poids d'un point ne représente que le quart du poids de la valeur située près du centre de la distribution.

Dans le doute au sujet de la position de la ligne, la prudence dicte de diminuer sa pente, ce qui implique une plus grande variation. À mesure que la pente diminue, l'intervalle de confiance de la CE 50 s'élargit.

Une fois la ligne ajustée, il est très simple d'en noter le point d'intersection avec l'effet à 50 % et, de ce point, de descendre jusqu'à la CE 50 sur l'axe des concentrations.

27. La plupart des laboratoires possèdent un programme d'assurance qualité permettant l'examen indépendant des données, qui devrait déceler toute erreur de saisie des données. Si une vérification rigoureuse peut remplir une fonction du graphique tracé à la main, elle ne le remplace pas. Les programmes informatiques peuvent produire des estimations particulières à partir de certains ensembles de données et nous, opérateurs humains, avons tendance à accepter le résultat sans autre examen. Un laboratoire d'Environnement Canada a signalé un écart important entre le graphique tracé à la main et le résultat d'un programme acquis depuis peu. Le programme (et non la saisie des données) s'est révélé la cause du problème (d'après K.G. Doe, Environnement Canada, Moncton).

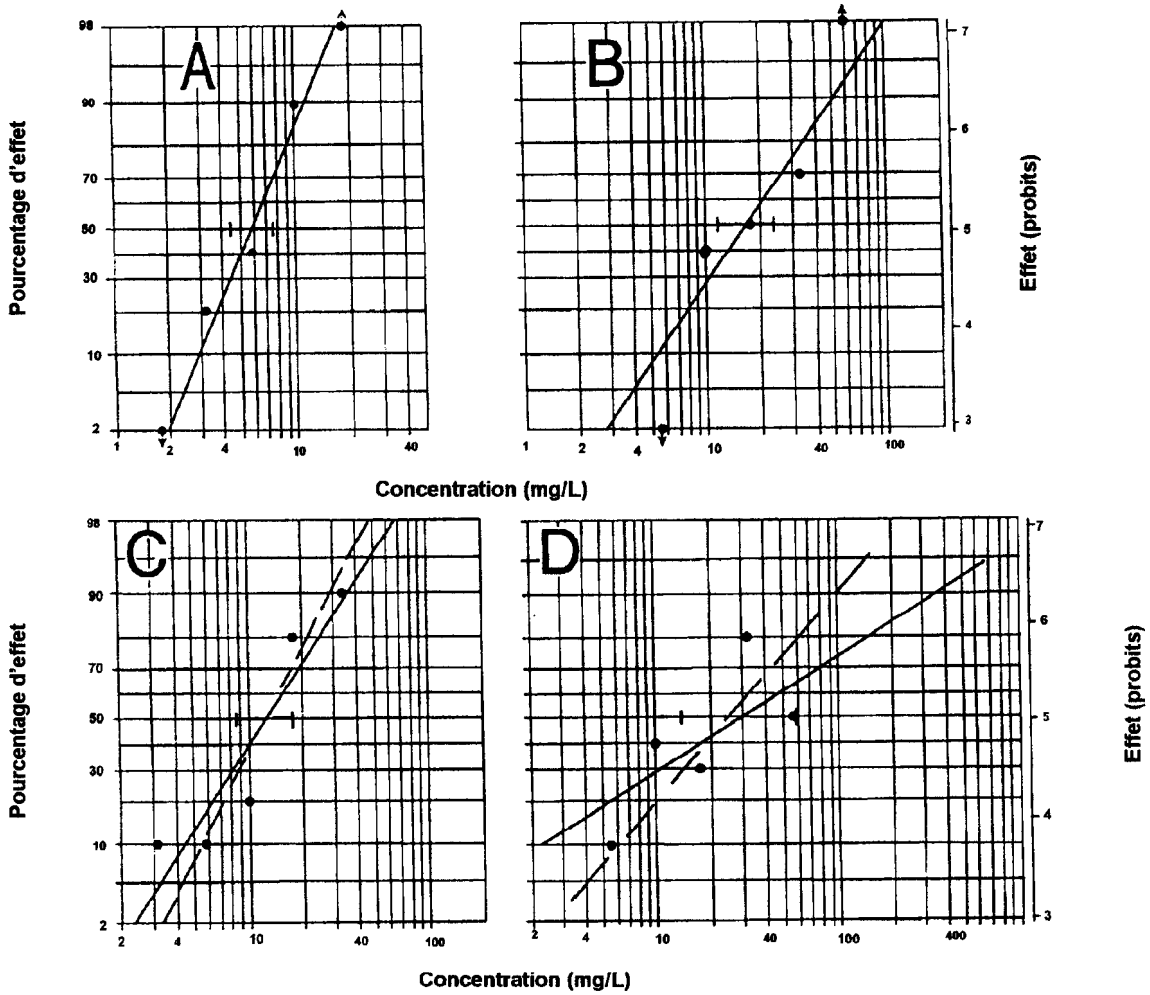


Figure 5. — Ajustement des droites des probits à vue d'œil à des ensembles représentatifs de données. Les graphiques A à D montrent les mêmes données que dans les exemples A à D du tableau 2. Les tiretés des graphiques C et D seraient des ajustements convenables, mais on préférerait les lignes ininterrompues, qui sont proches de celles que l'on calcule par régression probit. On montre, sous la forme d'un trait horizontal, les limites de confiance au seuil de 95 %, calculées par régression probit. Pour plus de précisions, voir le texte, particulièrement pour ce qui concerne le choix des droites dans certains des graphiques.

Les droites qui pourraient être ajustées à vue d'œil aux exemples de la fig. 5 sont discutées dans les exemples ci-dessous. Il sera commode de faire des comparaisons avec des lignes calculées par les méthodes statistiques formelles de régression probit, même si les méthodes mathématiques sont traitées dans le § 4.5.

Exemple A (fig. 5A). — L'emplacement de la droite ajustée est évident. Les données observées s'alignent bien, et la plupart des gens placeraient la droite très près de celle que l'on voit. Cette droite est essentiellement celle que l'on calcule par la régression probit, et son ajustement est bon, puisque la valeur de khi-deux est relativement faible ($\chi^2 = 1{,}11$ [tableau 2]). Les limites calculées de confiance au seuil de 95 % proviennent du tableau 2. L'intervalle de confiance est étroit, comme on s'y attendrait avec un ensemble cohérent de données et une droite des probits à la pente accentuée.

Exemple B (fig. 5B). — L'expérimentateur pourrait très bien ajuster la droite montrée, qui est

essentiellement identique à celle que l'on estime par la régression probit informatisée. On pourrait être tenté d'augmenter la pente, pour se rapprocher des valeurs extrêmes à 0 et à 100 %. Voici toutefois un bon exemple du poids moindre qui est accordé aux valeurs extrêmes : les trois points centraux influent fortement sur la droite. Les deux valeurs extrêmes ont cependant un petit effet, sinon la ligne calculée aurait eu une pente moins grande pour frôler les trois points centraux.

Exemple C (fig. 5C). — La plupart des gens considéreraient probablement le tireté comme un ajustement convenable. Il arrive presque à traverser les groupes supérieur et inférieur de points et parvient presque à réduire au minimum les distances verticales entre la droite et les points. Le tireté ajusté à vue d'œil permettrait d'estimer à peu près la même CE 50 que celle que l'on a calculée par régression probit à l'ordinateur (droite continue). Il pourrait sembler curieux que la droite des probits calculée se situe à la droite des deux points supérieurs. Apparemment, les calculs ont diminué la pente pour que la droite s'ajuste à la tendance globale de tous les points, reconnaissant la variation appréciable de cet ensemble de données (χ^2 relativement élevé de 3,5 [tableau 2]).

Exemple D (fig. 5D). — Des données aussi variables que celles-là pourraient être observées à la suite d'essais. Le tireté pourrait être un choix raisonnable pour une droite ajustée à vue d'œil. Il est moins que parfait en accordant un poids excessif à la valeur de 10 %, peu importante, et laissant une grande distance verticale au-dessus de la valeur de 50 %, à droite, alors que cette valeur possède, de fait, le poids maximal. Néanmoins, cette droite potentielle ne surestime que légèrement la CE 50 estimée par régression probit informatisée. Cette droite calculée possède une pente moins grande, en partie pour tenir compte de la plus grande influence des trois points centraux. La pente moins grande est également le signe d'une plus grande variation, comme en témoigne le khi-deux élevé de 5,5.

La conclusion générale à tirer de ces exemples est que les calculs statistiques et une ligne ajustée à vue d'œil conduisent souvent à des estimations similaires de la CE 50. Une autre conclusion manifeste est qu'une droite des probits de données variables, bien estimée, pourrait avoir une pente moins grande que celle que l'on obtient par un ajustement à vue d'œil.

4.2.4 Effets chez les organismes témoins

Repères

- *La plupart des méthodes d'Environnement Canada relatives à des essais quantiques tolèrent chez les organismes témoins des effets de pas plus de 10 % (≤ 10 %), bien que certains essais tolèrent un taux d'effet de 30 % chez des espèces particulières. Aucun facteur de correction n'est appliqué à l'effet qui, chez les organismes témoins, ne dépasse pas les limites tolérables, mais des effets plus grands invalident l'essai. Il faudrait en chercher la ou les causes et, si possible, recommencer l'essai.*
- *Dans le cas particulier de l'essai de toxicité sublétale quantique d'Environnement Canada employant des œufs de salmonidés, on applique une correction, par la formule d'Abbott, pour tenir compte des œufs non fécondés au début de l'essai. Cette correction est satisfaisante parce que la fécondation survient avant l'ajout du toxique. Une correction quelque peu semblable est effectuée dans l'essai de fécondation des oursins.*
- *Les programmes informatiques commerciaux pourraient ne pas suivre l'approche d'Environnement Canada concernant les effets observés chez les témoins, de sorte que l'expérimentateur doit comprendre le fonctionnement du programme.*
- *Dans le cas de la recherche ou des autres essais non assujettis aux méthodes d'Environnement Canada, la meilleure façon de se tirer d'affaire avec un effet révélé par les données quantiques chez les témoins, consiste à effectuer une analyse par un programme informatique qui effectue les estimations du maximum de vraisemblance du paramètre de mortalité chez les témoins. Sinon, on pourrait corriger l'effet observé chez les témoins par la formule d'Abbott, mais la méthode souffre de problèmes conceptuels de base, des points de vue biologique et statistique. Dans le cas inhabituel où l'effet exercé chez les témoins dépasse l'effet observé à une concentration donnée, cette formule donne une réponse curieuse, et on devrait toujours faire la correction à 0 %.*

Parfois, un effet à 10 % pourrait se manifester chez les organismes témoins, même dans des conditions favorables. Cela n'invaliderait pas les essais, et aucune correction ne devrait être appliquée pour un effet de cette amplitude. Certaines méthodes quantiques publiées par Environnement Canada spécifient que l'essai est invalide si les témoins manifestent un effet de plus de 10 % ; cela s'applique à la truite arc-en-ciel, à la daphnie (*Daphnia*) [EC, 1990c ; 1990d] et à plusieurs autres méthodes. Dans le cas des autres méthodes d'essai employant des organismes plus difficiles à garder au laboratoire, les taux de mortalité peuvent être plus grands dans des conditions qui semblent bonnes. Environnement Canada tolère des taux de mortalité, chez les témoins, de 20 % dans les essais universels avec les larves de têtes-de-boule et de 30 % dans le cas des essais de référence avec certains amphipodes (EC, 1992b ; 1998b).

Dans le cas des essais quantiques de toxicité aiguë d'Environnement Canada, les méthodes habituelles d'analyse statistique ne donnent aucune option permettant de corriger l'effet observé chez les témoins (par ex. EC, 1990a, b, c). [Une estimation du maximum de vraisemblance permettrait de tenir compte de l'effet observé chez les témoins, mais on y recourt actuellement peu de façon systématique.] Avec les méthodes habituelles d'analyse, l'essai serait simplement invalidé si l'effet observé chez les témoins dépassait la limite spécifiée dans les consignes. Les résultats seraient rejetés, et l'essai pourrait être répété, si on le voulait (et si cela était possible).

Même si un l'effet observé chez les témoins est acceptable selon la méthode d'Environnement Canada, on peut soupçonner que quelque chose cloche dans les conditions expérimentales ou dans l'état de santé des organismes. On devrait en chercher la cause apparente et, si on la trouve, on devrait tenter de la supprimer. Tout laboratoire qui a constamment éprouvé de grands effets chez les témoins ferait bien d'intensifier ses efforts afin de remédier au problème.

Essai de toxicité sublétale avec des œufs de salmonidés. — Cet essai (EC, 1998a) est un cas particulier pour la correction des effets observés chez les témoins. Dans ses préparatifs, une forte proportion des œufs peut ne pas être fécondée, mais l'expérimentateur ne peut reconnaître ces œufs que plus tard. Cette absence de fécondation ne peut pas interagir avec le toxique, toutefois, parce que le toxique est ajouté *après* la fin de la fécondation. Il n'y a aucune raison, si ce n'est le toxique, pour s'attendre à ce que les œufs, une fois fécondés, n'évoluent pas normalement et dans une proportion normale. Autrement dit, il ne peut pas y avoir d'interaction physiologique entre la réussite de la fécondation initiale et l'action du toxique. Dans ce cas particulier, on peut appliquer une correction pour tenir compte des œufs non fécondés, en employant la formule d'Abbott, décrite dans le texte qui suit. Dans ces circonstances, certains des problèmes conceptuels majeurs découlant de la correction d'Abbott ne s'appliquent pas. C'est pourquoi Environnement Canada recommande la formule d'Abbott pour cet essai avec des salmonidés, pour toute proportion raisonnable d'œufs non fécondés chez les témoins, notamment à de faibles taux de 10 % et moins. Après correction, les écarts entre les concentrations expérimentales et les témoins, dans les proportions d'œufs qui ne se développent pas, sont imputés à l'action de la matière à l'étude.

Dans l'essai de fécondation des oursins (EC, 1992f), on utilise une formule équivalente à la correction d'Abbott, dans l'analyse pour déterminer la CI p, marche à suivre usuelle avec cet essai de toxicité.

Programmes informatiques. — Les programmes disponibles ne suivent pas nécessairement les approches d'Environnement Canada à l'égard de l'effet observé chez les témoins. Certains programmes pourraient utiliser des méthodes sophistiquées d'estimation du maximum de vraisemblance pour estimer l'effet « réel » du toxique, sans l'effet observé chez les témoins (§ 4.5.5). L'effet observé chez les témoins devrait toujours se situer à l'intérieur des limites indiquées, si l'on veut utiliser les résultats de l'essai sous l'égide d'Environnement Canada. D'autres programmes informatiques pourraient automatiquement appliquer la formule d'Abbott, ce qui ne conviendrait pas à la plupart des méthodes publiées par Environnement Canada.

En conséquence, l'expérimentateur *doit* comprendre exactement comment tel programme informatique fonctionne à l'égard des effets observés chez les témoins. (Les programmes informatiques sont l'objet d'une discussion dans les alinéas qui suivent.) le

programme de Stephan (Stephan *et al.,* 1978) et certaines de ses adaptations n'accepte aucun effet observé chez les témoins. On peut recourir aux programmes TOXSTAT 3.5 et CETIS (v. ces noms sous la rubrique « Références ») pour corriger (ou non) l'effet observé chez les témoins. Le programme TOXCALC 5.0 applique la formule d'Abbott dans la régression probit, lorsqu'il le juge à propos. Choisir un programme approprié est la meilleure façon d'éviter de voir appliquer par l'ordinateur une correction non souhaitée à l'effet exercé chez les témoins. (Quoi qu'il en soit, l'effet observé chez les témoins devrait se situer dans les limites spécifiées dans la méthode d'essai d'Environnement Canada).

Utilisation de méthode du maximum de vraisemblance. — La meilleure façon de s'en tirer avec les effets observés chez les témoins est d'utiliser un progiciel qui utilise la méthode du maximum de vraisemblance (MMV ; v. le § 4.5.5). Les programmes offrant la MMV estiment deux paramètres pour décrire le modèle adopté et un troisième pour l'effet observé chez les témoins. Le paramètre de toxicité tel que la CE 50 est estimé pour l'effet du toxique *uniquement,* sans tenir compte de l'effet observé chez les témoins.

La MMV est offerte depuis longtemps dans les principaux progiciels tels que SAS (1988 ; 2000). Ces progiciels importants de statistique pourraient ne pas être accessibles dans tous les laboratoires. Les logiciels habituels pour les essais de toxicité (au moment d'écrire ces lignes, il s'agit notamment de CETIS, de TOXCALC et de TOXSTAT) se fondent sur la méthode classique des « moindres carrés itérativement repondérés ».

Même une méthode sophistiquée (le vrai maximum de vraisemblance effectué par SAS) ne fonctionne que dans les limites du test particulier. Le modèle sépare l'effet observé chez les témoins, mais *il ne compense pas la modification globale de résistance chez les organismes en expérience,* si cette modification a été provoquée par la maladie ou par quelque facteur semblable. En clair, la CE 50 pourrait dénoter des organismes affaiblis, peu résistants au toxique. Aucun modèle simple ni méthode simple de modélisation ne saisit actuellement l'interaction entre l'effet du facteur « milieu » et l'effet du toxique [28] (v. la discussion qui suit sur la formule d'Abbott). Le remède consiste à exécuter les essais dans de bonnes conditions, avec un bassin d'organismes en bonne santé.

Limites de la formule d'Abbott. — Cette méthode (Tattersfield et Morris, 1924 ; Abbott, 1925) mathématique simple permet de corriger les effets observés chez les témoins. Certains exemples de corrections sont présentés dans le tableau 1 et la fig. 6. La formule est donnée par l'équation 2. À noter que l'on y utilise des proportions, par ex. 3 organismes sur 10 sont saisis sous la forme fractionnaire 0,3.

$$P = \frac{P^* - C}{1 - C} \qquad (2)$$

Où :

P = la proportion corrigée d'organismes manifestant l'effet ;
P* = la proportion observée d'organismes manifestant l'effet ;
C = la proportion d'organismes témoins manifestant l'effet.

La formule d'Abbott se fonde sur l'hypothèse peu probable que l'effet constaté chez le témoin est tout à fait distinct de l'effet du toxique et n'influe pas sur lui. Comme les faits ont montré que l'hypothèse ne tient pas (compte rendu dans Hewlett et Plackett, 1979), la formule introduit une correction biaisée. Dans la situation où le témoin subit un grand effet, il pourrait y avoir combinaison de l'effet dû au toxique et de quelque(s) autre(s) facteur(s) ayant causé l'effet du milieu. Par exemple, les organismes affaiblis par une mauvaise alimentation pourraient être moins résistants au toxique, ce qui entraînerait l'estimation d'une CE 50 inférieure, dans leur cas, à la CE 50 correspondante estimée pour des organismes bien nourris [29]. Comme

28. Le Dr W. Slob (Institut national de santé publique et de l'environnement, des Pays-Bas, 2003, communication personnelle) signale qu'une telle méthode fait partie du logiciel PROAST, que d'autres chercheurs sont en train de développer.

29. D'autres problèmes sont imputables à la formule d'Abbott. Celle-ci corrige le nombre d'organismes réagissant, mais non le nombre d'organismes éprouvés. L'effet observé chez les témoins est considéré comme une constante, et il n'est tenu nullement compte de l'incertitude qui lui est associée (sa variance). Cet omission dans l'estimation de la CE 50 provoque une

Tableau 1. — Exemples de corrections apportées par la formule d'Abbott à divers effets observés chez les témoins dans un essai de toxicité quantique. Les résultats hypothétiques sont semblables à ceux de l'exemple B du tableau 2, mais les concentrations faibles et fortes sont moins extrêmes. Dans les quatre colonnes de droite, l'effet observé chez les témoins a été remplacé par un effet de 0, 10, 20 et 30 %. Les résultats correspondant à chaque concentration expérimentale ont été corrigés par la formule d'Abbot pour tenir compte de ces effets observés chez les témoins. On a ensuite appliqué une régression probit (Stephan *et al.*, 1978) pour calculer les résultats qui figurent dans les quatre rangées inférieures et qui sont représentés dans la fig. 6.

Concentration (quantité/litre)	Nbre d'organismes testés	Nombre d'organismes touchés, corrigé à la main, pour un effet, observé chez les témoins de...			
		0 %	10 %	20 %	30 %
56	10	8	7,78	7,50	7,14
32	10	7	6,67	6,25	5,71
18	10	5	4,44	3,75	2,86
10	10	4	3,33	2,50	1,43
5,6	10	2	1,11	0	0
CE 50		16,5	20,3	25,2	30,1
Limites de confiance		7,85 ; 31,3	12,0 ; 31,9	16,9 ; 43,5	20,7 ; 53,5
Pente		1,65	1,89	2,38	2,53
Khi-deux		0,136	0,286	1,26	0,606

nous l'avons mentionné, aucune méthode connue ne permet de corriger les interactions possibles entre le toxique et le facteur, quel qu'il soit, de l'effet observé chez le témoin.

Nous examinons les effets de la formule d'Abbott à l'aide des données hypothétiques du tableau 1. Les corrections sont affichées dans les colonnes de ce tableau, pour des effets observés chez les témoins de plus en grands, allant de 0 à 30 %. À mesure que l'effet observé chez les témoins augmente, les résultats corrigés sont de plus en plus modifiés. La CE 50 estimée augmente de 80 %, la pente de plus de 50 %, mais les limites de confiance restent proportionnellement semblables. Khi-deux augmente, mais reste au moins 6 fois plus petit que la valeur critique. Ces modifications sont représentées dans la fig. 6, où la droite des probits se déplace vers le bas et la droite dans les graphiques successifs et où la pente augmente. D'autres exemples pourraient se comporter différemment, mais la CE 50 et la pente augmenteraient presque toujours avec des corrections plus grandes des effets observés chez les témoins.

La formule d'Abbott et les effets grands ou anormaux observés chez les témoins. — Il faudrait comprendre que, chez les organismes que l'on peut garder au laboratoire, un effet observé de 20 %, 30 % ou plus jetterait un doute sérieux sur la validité et l'utilité de la CE p. En outre, la correction de l'effet

sous-estimation de la variabilité de ce paramètre de toxicité. Si on utilise la régression probit, l'hypothèse de la linéarité entre les probits et le logarithme de la concentration ne tient plus lorsqu'on observe un effet chez les témoins. Si on applique la correction à plusieurs concentrations, cela introduit une corrélation entre les concentrations, bien que celles-ci doivent être indépendantes. Si on observe à une certaine concentration un effet à 100 %, la formule d'Abbott ne modifie pas cet effet, c'est-à-dire qu'elle attribue la totalité de l'effet au toxique, sans tenir compte de la cause de l'effet observé chez les témoins.

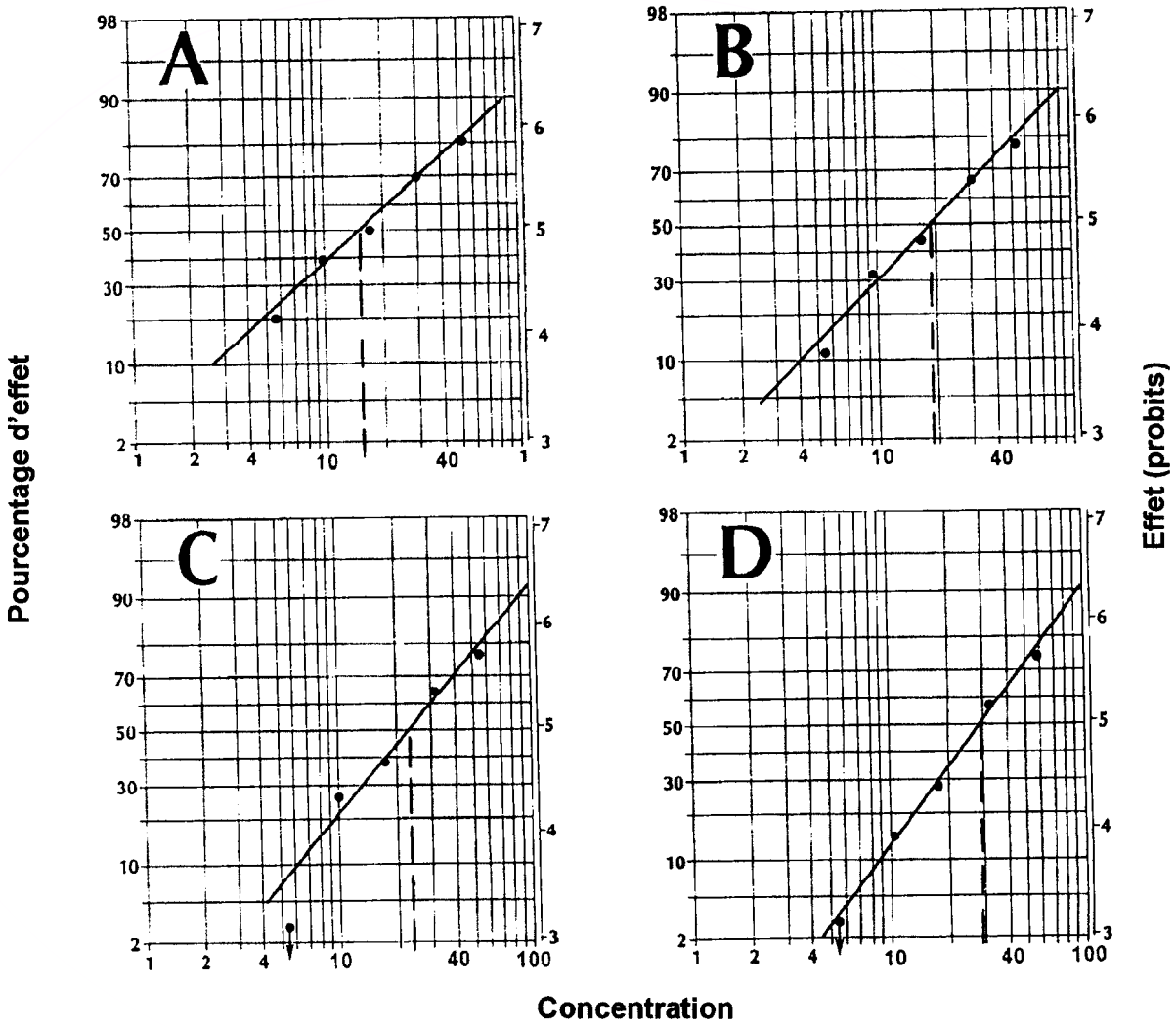


Figure 6. — Résultats des corrections apportées au moyen de la formule d'Abbott aux résultats d'un essai quantique, pour tenir compte de l'effet observé chez les témoins. Les droites représentent les valeurs calculées, présentées dans le tableau 1. Le graphique A est un exemple d'effet nul, observé chez les témoins. Dans les graphiques B, C et D, les mêmes résultats sont corrigés pour des effets observés chez les témoins de 10, 20 et 30 %, et la CE 50 estimée passe d'environ 16 à environ 30. La pente s'accentue également, parce que, avec des corrections plus importantes, les petits pourcentages d'effet (faibles concentrations) sont déplacés vers le bas, vers l'effet nul, dans une proportion plus grande que les grands pourcentages d'effet.

observé chez les témoins par la formule d'Abbott entraînerait les difficultés conceptuelles majeures que nous décrivons. Si ces difficultés étaient acceptées par un expérimentateur affranchi, pour quelque raison que ce soit, des exigences d'Environnement Canada, cette formule pourrait servir à corriger les effets en question jusqu'à un taux d'environ 30 %. La méthode a pour effet d'augmenter de façon appréciable la CE 50 estimée, comme le montrent la fig. 6 et le tableau 1.

Dans le cas de petits effets observés chez les témoins, d'au plus 10 %, il serait rarement souhaitable

d'appliquer une correction, peu importe le but de l'essai. Cet effet, phénomène accidentel, inhabituel ou aléatoire, pourrait avoir peu influé sur la CE 50 de la matière toxique étudiée. Une « correction » empirerait alors l'estimation de la CE 50.

Si l'effet observé chez les témoins était supérieur à l'effet observé à une concentration donnée, la formule d'Abbott donnerait une réponse curieuse. L'effet observé serait corrigé de façon à égaler une valeur négative, ce qui n'est pas logique, puisque cela implique que l'*on compterait plus d'organismes non touchés que d'organismes effectivement soumis à cette concentration.* Finney (1971) a recommandé d'utiliser les probits correspondants pour la valeur négative et de poursuivre les calculs, puisque cela est simplement une variation due à l'échantillonnage. Cependant, l'expérimentateur risquerait ne plus pouvoir maîtriser le programme informatique qu'il utilise de cette façon. Certains programmes se sont révélés effectuer la correction à une valeur négative, puis ignorer le signe négatif, utiliser la valeur positive pour créer un probit utilisable dans l'analyse et poursuivre avec un calcul faux de la CE 50 !

Nous recommandons donc ceci : *S'il a été décidé d'appliquer des corrections à l'effet observé chez les témoins et si cet effet égale ou excède un effet observé et qu'il est incertain que le programme informatique puisse manipuler une valeur négative de l'effet, alors :* a*) corriger tous les effets observés par une méthode manuelle ;* b*) attribuer à l'effet aberrant la valeur de 0 % plutôt qu'une valeur négative ;* c*) saisir les effets corrigés sans la valeur attribuée au témoin.*

On peut voir un exemple dans la dernière colonne du tableau 1. Pour des effets de 30 % et de 20 % observés chez les témoins à la concentration 5,6, la formule d'Abbott corrigerait les valeurs à − 0,143 ou à − 1,43 organisme. Nous avons plutôt inscrit 0.

Hubert (1984) déclare que la « formule d'Abbott ne s'applique qu'aux taux de mortalité qui excèdent l'estimation du taux naturel de mortalité », mais cela ne semble pas raisonnable. Si un effet observé était inférieur ou égal à l'effet observé chez les témoins, on n'y toucherait pas, et on l'imputerait au toxique.

Il est clair que, en ce qui concerne la formule d'Abbott, l'expérimentateur doit choisir un programme informatique approprié et comprendre exactement comment ce programme traite les effets observés chez les témoins. Parmi les programmes usuels au moment d'écrire ces lignes, TOXCALC 5.0 applique la formule d'Abbott dans les situations appropriées, tandis que TOXSTAT 3.5 et CETIS le font dans le cadre de plusieurs options. Le programme de Stephan *et al.* (1978) et celui de l'OMEE (1995) posent, par hypothèse, que l'effet observé chez les témoins est nul. Si on appliquait la formule d'Abbott, une façon détournée mais certaine d'obtenir le résultat voulu d'un programme serait celle qui a été exposée : calculer à la main les corrections pour chaque concentration, puis saisir les versions corrigées comme si elles étaient les observations brutes. Ne saisir aucune donnée sur les témoins (ou inscrire un effet nul, si le programme exige la saisie de données relativement aux témoins). Les valeurs corrigées seraient probablement sous forme décimale (par ex. 3,33 vers de terre sur 10), mais la plupart des programmes de statistique acceptent « volontiers » de telles fractions.

4.2.5 Limites de confiance de la CE p

Repères

- *Il faut signaler les limites de confiance au seuil de 95 % de la CE 50 ; elles permettent d'estimer la variation interne des résultats de l'essai. Un rapport de 1,3 entre la CE 50 et la limite de confiance représente un intervalle de confiance étroit et une bonne précision, tandis qu'un rapport de 1,5 à 1,8 est courant et acceptable.*
- *Les limites de confiance ne renseignent que sur la variation d'un essai particulier. Elles ne disent rien de la variation globale des essais avec un toxique donné.*
- *Il est bon de signaler également la pente de la droite concentration-effet, qui permet de recréer cette droite ultérieurement, si on le souhaite. On devrait également signaler le khi-deux de la qualité de l'ajustement.*
- *L'intervalle de confiance est le plus étroit à la CE 50 et il s'élargit aux taux supérieurs et inférieurs d'effet.*
- *À cause de la variabilité, il y a des inconvénients à choisir les paramètres de toxicité aux faibles*

*taux d'effet, comme la CE 20. Le taux choisi d'effet (*p *% dans la CE p) ne devrait jamais être dans la région où l'on prévoit que l'effet chez les témoins sera observé et il ne devrait probablement jamais être inférieur à 10 %* (*CE 10*).

Les résultats communiqués doivent toujours comprendre les limites de confiance au seuil de 95 % de la CE 50. La seule exception serait les résultats d'essais qui n'ont révélé d'effet partiel à aucune concentration. La fig. 4 montre que ces essais seraient analysés par la méthode binomiale, qui ne donne pas de limites de confiance. Il est également bon de signaler la pente de la droite ajustée de la relation concentration-effet et le résultat du khi-deux pour la qualité de l'ajustement. La connaissance de la pente permet de reconstituer la droite, si on le veut ; faute de pente, la description de la relation entre la concentration et l'effet est insuffisante.

L'expérimentateur doit toujours se rappeler que les limites de confiance d'un seul essai de toxicité ne révèle que le degré de précision interne de cet essai, quel que soit le nombre d'organismes utilisés, dans les conditions qui existaient au moment de l'essai et avec les incertitudes associées au modèle. *Ces limites ne doivent pas être confondues avec les limites globales de la CE 50 d'une matière donnée.* Les estimations de la CE 50 peuvent différer considérablement d'un moment à autre et d'un lieu à l'autre, relativement à la même espèce et dans des conditions semblables. Par exemple, si on souhaitait définir les limites probables de la toxicité d'un effluent particulier, les limites de confiance d'un essai de toxicité ne les révéleraient pas. Il faudrait analyser plusieurs échantillons de l'effluent. Ensuite, la variation du paramètre de toxicité estimé grâce à ces essais constituerait la base de la prévision des limites de la toxicité de l'effluent, dans les conditions qui existaient pendant la période d'échantillonnage. La variation est traitée dans le guide d'Environnement Canada sur l'interprétation des données environnementales (EC, 1999a).

Dans l'exemple A du tableau 2, la plupart des limites calculées de confiance (supérieures ou inférieures) diffèrent de la CE 50 par un facteur d'environ 1,3 — ce qui est une bonne précision dans un essai de toxicité en milieu aquatique. Dans les essais avec le poisson, les laboratoires trouvent souvent des facteurs de 1,3 à 1,5 entre la limite de confiance et la CE 50, en utilisant 10 poissons par concentration. L'expérience montre que, la plupart du temps, un facteur d'environ 1,8 signifierait une précision acceptable [30]. Pour ce qui concerne les données variables telles que celles de l'exemple D du tableau 2, les limites de confiance pourraient être extrêmes ; certaines des limites supérieures de confiance estimées par régression probit sont une dizaine de fois plus grandes que la CE 50. L'expérimentateur devrait être prêt à observer parfois de grands intervalles de confiance. Parfois, on peut améliorer les limites en choisissant un modèle mieux ajusté aux données, si les limites ne se conforment pas au modèle habituel. Si, malgré cela, les limites ne sont pas considérées comme satisfaisantes, il ne reste qu'à reprendre l'essai.

Parfois, les limites supérieure et inférieure pourraient sembler à peu près symétriques par rapport à la CE 50, sur une échelle logarithmique, mais, normalement, un certain degré d'asymétrie existerait (v. le texte qui suit et la fig. 7).

L'intervalle de confiance est commandé par la pente de la droite de la relation dose-effet (qui donne une idée de la variation), par la dispersion des points observée de part et d'autre de la droite et par le nombre d'organismes utilisés à chaque concentration. Si chaque organisme était affecté par des concentrations tout à fait différentes de toxique, la droite des probits aurait une faible pente, qui contribuerait à un intervalle étendu de confiance. Cela pourrait survenir en raison du mode d'action du toxique, sans que cela soit nécessairement le signe d'un vice du mode opératoire. Une petite pente pourrait, cependant, être causée par la négligence, par ex. l'acclimatation incomplète des poissons à l'eau de dilution (Calamari *et al.,* 1980).

On peut améliorer la précision de la CE p estimée en employant plus d'organismes dans l'essai, mais une amélioration importante exige souvent un nombre

30. Hodson *et al.* (1977) estiment qu'un essai typique de toxicité employant 10 poissons par concentration et trois concentrations causant des effets partiels aurait une limite supérieure de confiance près de 2,1 fois la valeur de la CE 50. Les exemple A, B et C du tableau 1 mentionnent des limites de confiance qui sont 1,3, 1,4 et 1,4 fois, respectivement, la valeur des CE 50. Les estimations de la variation par Hodson *et al.* (1977) semblent quelque peu plus grandes que ce que l'on constate habituellement dans beaucoup de laboratoires.

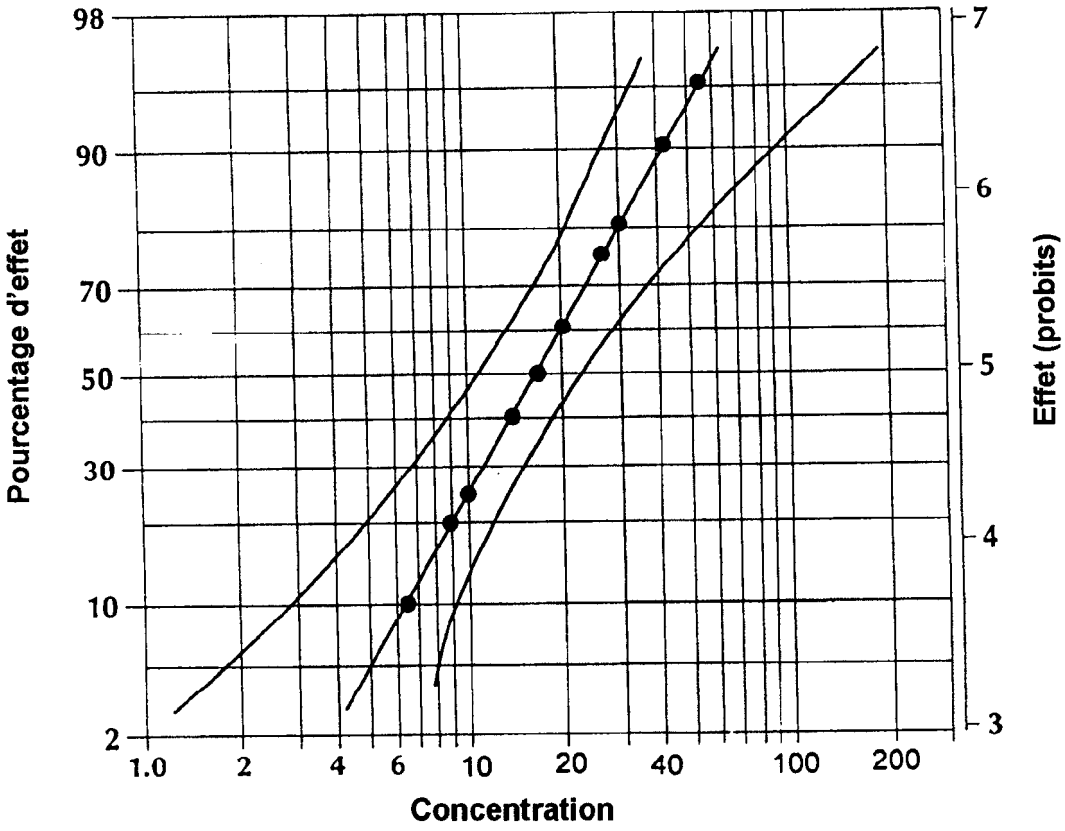


Figure 7. — Élargissement de l'intervalle de confiance des concentrations efficaces autres que la CE 50. La droite des probits et les limites de confiance sont tirées de l'exemple B du tableau 2. Les gros points noirs traversés par la droite des probits correspondent aux valeurs calculées des CE p ayant servi à tracer la droite et les limites de confiance. Les valeurs ont été estimées par le programme TOXCALC.

impraticable d'organismes, comme il en est question dans le § 2.5.

La fig. 7 représente des résultats réguliers et dont l'intervalle de confiance est assez étroit. (À noter que les données réelles ne sont représentées dans cette figure. Les points sont les valeurs calculées le long de la droite ajustée.) Les limites de confiance au seuil de 95 % de la CE 50 sont les concentrations de 11,9 et de 23,7, qui diffèrent de la CE 50 par un facteur d'environ 1,4, considéré comme satisfaisant pour un essai de toxicité (v. le texte qui précède).

La fig. 7 montre que l'intervalle de confiance au seuil de 95 % diffère considérablement selon les pourcentages différents d'effet, s'élargissant à mesure que l'on s'éloigne de la CE 50. Vers les valeurs extrêmes de la relation concentration-effet, les intervalles sont très étendus. Cela montre pourquoi l'effet médian est un bon choix comme paramètre de toxicité et pourquoi il n'est pas judicieux d'adopter comme paramètres de toxicité des effets très petits, par ex. la CE 10, qui a une résonance protectrice séduisante.

La fig. 7 montre aussi que les limites de confiance présentent une certaine asymétrie horizontale. C'est le cas normal. À l'origine, les limites sont calculées par rapport aux *effets observés* à des concentrations fixes, de sorte que, en tout point sur la droite des probits, elles sont verticalement symétriques par rapport à la droite (v. la discussion à ce sujet dans le § 9.4). Les estimations inversées donnent ensuite les limites de confiance *en concentrations,* conformément aux souhaits de l'expérimentateur. Ces limites sont toujours asymétriques, du moins légèrement, souvent à un degré

notable. Les limites correspondant aux extrémités de la distribution sont fortement asymétriques.

4.2.6 CE 20 ou autres concentrations que la CE 50

Dans les essais de toxicité quantique, on a pour habitude d'estimer un effet médian (CE 50), parce que cette concentration représente l'organisme « médian » ou « typique » et parce qu'on lui associe l'intervalle de confiance le plus étroit, c'est-à-dire la précision la plus grande. En même temps, il y a une forte demande de concentrations ou de paramètres perçus comme « plus protecteurs », c'est-à-dire associés à des effets proportionnels moindres, tels que la CE 20 ou la CE 25. Une façon de ménager la chèvre et le chou est d'accepter la CE 50 et sa grande précision, puis d'appliquer un facteur approprié pour obtenir une concentration qui s'appliquerait à une fraction plus petite de la population d'organismes. Cela a des qualités et des défauts. L'approche plus directe, qui emploie les mêmes méthodes générales que pour la CE 50, consiste à estimer la CE 20 (ou la CE du pourcentage voulu [la CE x]) directement et de tolérer un intervalle plus large de confiance.

Il faudrait veiller cependant à ne pas tenter d'estimer une valeur correspondant à un pourcentage d'effet très faible. Si la CE 01 peut sembler tentante en tant que concentration ayant un effet négligeable, elle correspond à d'énormes difficultés conceptuelles, et la variabilité de l'estimation la rend très peu fiable (fig. 7). Toute tentative d'estimation d'une CE p qui serait semblable aux éventuels effets observés chez les témoins serait d'une validité et d'une signification contestables. La règle suivante semblerait raisonnable : *ne jamais tenter d'estimer un paramètre de toxicité qui se situe dans l'intervalle acceptable d'effet chez le ou les témoins.* Outre cela, toute valeur de *p* serait suspecte si elle était inférieure au pourcentage minimal d'effet observé aux concentrations d'essai. Ainsi, la valeur minimale acceptable de *p* dépendrait des résultats de l'expérience. Ce pourrait être de moins de 10 %, dans le cas d'une expérience de très grande envergure ou ce pourrait être de 20 % ou même plus dans une autre expérience.

Le progrès dans l'estimation de la CE x aux petites valeurs de *x,* présente un intérêt considérable pour l'estimation des concentrations « inoffensives » ou « sans effet » des contaminants pour l'homme comme pour les systèmes naturels. Noppert *et al.* (1994) ont étudié cela sous l'impulsion de l'OCDE et ils ont conclu que la meilleure approche serait de modéliser la CE x, plutôt que d'employer une technique de test d'hypothèse. Cependant, ils ont fini par proposer 5 ou 10 % comme valeur préférable de *x,* plutôt qu'une valeur plus près de 0. Les techniques de régression permettant d'estimer les faibles valeurs de CE x se sont également révélées l'approche supérieure, d'après les conclusions de Moore et Caux (1997).

Il serait particulièrement futile de tenter d'estimer une concentration qui manquerait de peu de nuire à un seul organisme (la « CE 00 »). On ne peut pas l'estimer explicitement, parce qu'elle dépendrait de la taille de population (un organisme sur cent ? sur mille ? sur un million ?). De plus, aucune méthode statistique n'est conçue pour arriver à un tel un tel paramètre de toxicité. (Cependant, les § 5.2 et 5.3 font allusion à des techniques de modélisation plus sophistiquées qui, par extrapolation à partir des résultats d'essais de toxicité aiguë, permettent d'arriver à des seuils d'effets chroniques.)

Les restrictions s'appliquant aux types de données se prêtant à l'estimation de CE p « non médianes » seraient celles qui sont énumérées au début des § 4.2 et 4.2.1. Il faudrait remplacer *p* par la valeur appropriée ; par ex. l'analyse pourrait exiger un effet égal ou supérieur à 20 % plutôt qu'à 50 %.

Plusieurs programmes informatiques permettent d'estimer les CE p non médianes par régression probit ou logit. Le gros progiciel SAS le fait, et le programme SPSS imprime une sélection de CE p englobées dans tout l'intervalle utile. CETIS, TOXCALC et TOXSTAT font de même ou on peut leur demander de le faire. (On peut trouver ces progiciels de statistique sous leurs noms, sous la rubrique « Références ».) Dans leur application de la méthode de Spearman-Kärber, ces progiciels n'estiment que la CE 50. Le programme de Stephan *et al.* (1978) et ses adaptations (OMEE, 1995) se bornent également à estimer la CE 50.

L'autre méthode d'estimation de faibles valeurs de paramètres de toxicité serait de commencer par l'estimation du paramètre médian, dont la précision est

plus grande (comme il est montré dans le § 4.2.5). On pourrait appliquer ensuite un facteur pour estimer une concentration censée causer le petit effet partiel auquel on s'intéresse, peut-être une concentration mal définie dans les résultats d'un essai donné de toxicité. Par exemple, on pourrait appliquer un facteur à la CE 50 pour aboutir à une CE 20 ou même à une CE 5 prévue. On pourrait choisir le facteur d'après la pente habituelle des droites de probits ou de logits obtenues à la faveur de tels essais. (Cette approche a, de fait, été utilisée pendant des décennies pour extrapoler les concentrations létales médianes à des concentrations censées être inoffensives et que l'on a utilisées comme objectifs de qualité de l'eau. Ce sont les « facteurs d'application » décrits dans EC, 1999a.) L'emploi de ces facteurs a l'avantage de pouvoir servir de point de départ relativement bien défini. Il a aussi le défaut d'être plus ou moins hypothétique selon le degré d'extrapolation.

4.3 Choix de méthodes

Repères

- *La régression probit ou logit par régression du maximum de vraisemblance (méthode du maximum de vraisemblance [MMV]) est la méthode usuelle, préférée pour les effets quantiques à trois concentrations ou plus, y compris deux concentrations exerçant des effets partiels. Le second choix va à la méthode communément utilisée de régression probit (ou logit) itérative, qui donne des estimations comparables à celles de la MMV. On recommande actuellement les régressions probit ou logit pour leur utilisation en routine, en raison de leur disponibilité et de leur commodité.*
- *Certains essais pourraient ne donner qu'un effet partiel, qui ne se prête pas à la régression probit ou logit. Pour ces ensembles de données, la méthode recommandée est celle de Spearman-Kärber. On devrait l'appliquer sans équeutage des données et, également, avec équeutage minimal (de 35 % à la limite).*
- *Si les concentrations successives donnent des effets à 0 et à 100 %, sans effet partiel, on devrait estimer la CE 50 approximative par la méthode binomiale. Cette dernière devrait aussi être utilisée si l'on obtient des résultats anormaux avec la méthode de Spearman-Kärber. La méthode binomiale ne donne pas de limites de confiance à 95 %, mais, plutôt, elle estime des limites prudentes à l'intérieur desquelles devrait se trouver la CE 50.*
- *La méthode de la moyenne mobile est valide, mais elle possède à l'égard des données les mêmes exigences que la régression probit ou logit, que nous recommandons plutôt.*
- *Pour l'analyse, on dispose de divers logiciels commerciaux et de logiciels de l'administration publique. L'utilisateur doit comprendre tout à fait les opérations utilisées par le logiciel qu'il choisit. Certains logiciels possèdent des inconvénients pour les besoins d'Environnement Canada ou nécessitent l'apport de données inutiles, conçues cependant pour les besoins d'organismes étrangers de réglementation.*

Nous recommandons les méthodes suivantes d'analyse pour les essais effectués dans le cadre de programmes élaborés par Environnement Canada. Les méthodes les plus souhaitables (1) et (2) ne conviendront pas à la plupart des données obtenues par des essais en routine, parce qu'elles exigent deux effets partiels. D'autres méthodes d'analyse figurent dans la liste, pour d'autres types de données. Les diverses méthodes acceptables sont décrites plus en détail dans le § 4.5.

1. *Régression probit ou logit par la méthode du maximum de vraisemblance* (§ 4.5.3). — On sait que cette méthode est offerte dans le progiciel de statistique SAS (1996). Elle a l'avantage d'être une méthode non biaisée, de prise en compte de l'effet observé chez les témoins et d'estimation d'un paramètre de toxicité fondé uniquement sur l'effet du toxique. Les calculs exigent deux effets partiels dans les données employées.
2. *Régression probit ou logit par itération*. — Les programmes utilisent la régression itérativement repondérée pour arriver à une estimation définitive. Les programmes informatiques les plus facilement accessibles suivent cette technique itérative « classique ».

Elle procure une analyse satisfaisante, parvenant à une solution équivalant à une estimation du maximum de vraisemblance. Cette méthode exige deux effets partiels.

3. *Méthode de Spearman-Kärber.* — Cette méthode n'est recommandée que si les résultats ne se prêtent pas à l'analyse par les deux méthodes susmentionnées. Les données doivent renfermer un effet partiel, plus un effet nul et un effet à 100 % ou des valeurs près de ces extrêmes. L'analyse devrait être effectuée sur des données : *a*) non équeutées ; *b*) ayant subi un équeutage « automatique » ou « minimal » d'au plus 35 %. Par inspection des résultats bruts et de leur représentation graphique, on devrait choisir le paramètre de toxicité le plus « raisonnable » des deux valeurs estimées. Si aucun n'est raisonnable, on devrait employer la méthode binomiale.

4. *Méthode binomiale.* — Cette méthode est à employer dans les cas où on ne possède aucun effet partiel, mais des effets à 100 % et à 0 %. Cette méthode serait également adoptée dans d'autres situations où on ne pourrait pas se servir des méthodes 1 à 3. Par exemple, on l'utiliserait s'il y avait un effet partiel et que la méthode de Spearman-Kärber avait donné des résultats anormaux, faute d'un effet à 0 et/ou à 100 % ou pour d'autres raisons.

5. *Méthode de la moyenne mobile.* — Le programme disponible pour cette méthode exige deux effets partiels. Il pourrait être utile, dans les situations inhabituelles dans lesquelles l'analyse par la méthode des probits ou des logits a échoué. Elle ne semblerait pas offrir d'avantage particulier dans d'autres situations.

6. *Méthode graphique de Litchfield-Wilcoxon.* — Elle n'est pas recommandée pour les rapports définitifs. Elle est utile pour vérifier les estimations informatisées, pour le travail de terrain ou pour les besoins de la formation.

La méthode la plus souhaitable d'estimation de la CE 50 est celle de l'**estimation** vraie **du maximum de vraisemblance** utilisant les probits ou les logits (MMV, méthode 1 susmentionnée). Elle repose sur l'hypothèse selon laquelle, à chaque concentration, une proportion des organismes en expérience sera touchée. Elle suppose en outre que ces proportions sont corrélées dans une courbe cumulative de fréquence qui passe d'un effet nul, aux faibles concentrations, à un effet à 100 %, aux fortes concentrations. La MMV tente d'estimer les valeurs des paramètres dans la relation, qui aboutiraient à la probabilité maximale d'observer les données effectivement collectées (v. le § 4.5.5). Une fois définie, la relation mathématique permet *de prévoir* la concentration censée produire un effet donné. La MMV peut être réalisée par le gros progiciel de statistique SAS, qui pourrait ne pas être disponible dans certains laboratoires ou ne pas être facilement utilisé par les expérimentateurs.

La régression probit itérative (méthode 2, ci-dessus) est offerte dans les grosses bibliothèques de logiciels faisant autorité, notamment les programmes SPSS et SYSTAT (énumérés sous leur nom sous la rubrique « Références ») et dans la plupart des autres progiciels commerciaux de toxicologie. En raison de sa disponibilité universelle, nous désignons la régression probit ou logit par itération comme la méthode usuelle pour une utilisation en routine. Plus loin (§ 4.5.1 à 4.5.6), nous abordons plus en détail les démarches méthodologiques et le choix des logits ou des probits.

Le « programme de Stephan » (Stephan *et al.*, 1978), bien connu, comprend la régression probit (méthode 2), la moyenne mobile (méthode 5), et c'est la seule source pratique de la méthode binomiale (méthode 4). Il a été mis au point par le Dr Charles E. Stephan et ses collègues de l'USEPA à Duluth (Minnesota) et il est utilisé depuis plus de deux décennies. On le recommande dans beaucoup de méthodes d'Environnement Canada, il a généralement été utilisé dans les laboratoires canadiens et il a été rendu accessible par le personnel de ces derniers. Le programme de Stephan a été adapté sous diverses formes. Une adaptation, au laboratoire d'Etobicoke du ministère de l'Environnement et de l'Énergie de l'Ontario, écrite par le Dr Gary F. Westlake, fonctionne sur une des premières plate-formes Windows (OMEE, 1995) ; dotée des méthodes des probits, de Spearman-Kärber et de la moyenne mobile, elle produit un graphique des résultats (nous l'appelleront ci-après *programme de l'OMEE*).

Parmi les divers programmes informatiques

commerciaux créés aux États-Unis, mentionnons CETIS et leurs prédécesseurs TOXSTAT 3.5 et TOXCALC 5.0. Ils peuvent analyser des données quantiques par diverses méthodes, mais, généralement, ils comprennent les méthodes des probits, des logits et de Spearman-Kärber. Un programme fiable en langage BASIC est décrit dans USEPA (1994a, annexe I ; 1994b, annexe H ; 1995, annexe H) et peut être obtenu de l'USEPA à Cincinnati (Ohio) ou sur le site Web http://www.epa.gov/nerleerd/stat2.htm. D'autres programmes ne pourraient pas convenir en raison de caractéristiques non appropriées au Canada [31].

Régression linéaire simple. — Ce type de régression, qui souffre de limites majeures, n'est pas recommandé. Il pourrait sembler une méthode mathématique évidente d'ajustement d'une droite aux données quantiques, telles que celles que montre la fig. 5, mais il n'est pas valide. En effet, la différence de valeur (« le poids ») entre les points, inversement proportionnelle à la variation, peut, à son tour, augmenter vers les extrémités supérieure et inférieure de la droite. Les poids doivent être intégrés dans le processus d'ajustement, mais il y a un hic : les pondérations ne peuvent être calculées qu'à partir de la droite ajustée et non des effets observés bruts (v. la note 32). Cela explique pourquoi on ne peut pas utiliser la régression simple et pourquoi il faut adopter des méthodes telles que l'itération. De temps en temps, des expérimentateurs naïfs utilisent erronément la régression simple en tentant d'estimer des CE 50.

4.4 Comparaison des estimations par diverses méthodes

Repères

- *La plupart des méthodes courantes d'analyse statistique des résultats des essais quantiques sont susceptibles de donner des estimations similaires de la CE 50 et de ses limites de confiance, si les données sont acceptables.*
- *On a analysé des exemples de bonnes données hypothétiques au moyen de diverses méthodes. On a obtenu des résultats semblables par les méthodes des probits, des logits, de Spearman-Kärber, de la moyenne mobile et de la transformation arc sinus, et elles concordent avec les résultats obtenus par ajustement d'une droite à vue d'œil. Les estimations de la CE 50 par les méthodes binomiale et de Gompertz ont été quelques peu plus élevées.*
- *Les limites de confiance étaient également semblables chez la plupart des méthodes, bien que la méthode de Spearman-Kärber avec équeutage aient donné un intervalle de confiance plus étendu. La méthode binomiale n'a pas donné de limites de confiance, mais, plutôt, un intervalle où se trouveraient les limites de confiance.*
- *Pour certains exemples avec seulement un effet partiel, la méthode de Spearman-Kärber sans équeutage des données a procuré de bonnes estimations des CE 50, tandis que la méthode avec équeutage n'a pas permis d'obtenir d'estimations. La méthode binomiale a également permis de bonnes estimations de la CE 50.*
- *Avec certains exemples de données irrégulières ou parmi lesquelles il ne se trouvait pas d'effet nul et total, la méthode de Spearman-Kärber sans équeutage a donné des estimations très aberrantes. Les estimations obtenues avec équeutage des données ont varié selon le type de données — certaines étaient excellentes et d'autres ont été améliorées, mais elles restaient divergentes. La méthode binomiale a échoué.*

31. L'USEPA peut influencer la conception des programmes informatiques commerciaux états-uniens, mais ces programmes pourraient ne pas être conformes aux pratiques d'Environnement Canada. La méthode de Spearman-Kärber a usuellement été offerte dans la procédure états-unienne, sans les limites recommandées ici pour Environnement Canada (§ 4.5.6). La méthode binomiale (et celle de la moyenne mobile) ne sont pas offertes dans les programmes états-uniens récents. À la place, ceux-ci offrent l'« interpolation linéaire » entre deux points de données (§ 4.5.9), et cela peut être satisfaisant et équivalent à la méthode binomiale s'il y a des effets nul et total successifs. L'expérimentateur devrait cependant s'assurer que l'interpolation linéaire a employé par défaut le logarithme des concentrations. Les programmes pourraient aussi exiger des renseignements qui ne sont pas appropriés au Canada, parce qu'ils formatent les résultats de façon à répondre aux exigences de l'USEPA en matière de rapports.

Dans le présent paragraphe, nous comparons les paramètres de toxicité quantique estimés au moyen de diverses méthodes statistiques. Dans le § 4.4.1, nous utilisons comme exemples des ensembles de données relativement bonnes. Dans le § 4.4.2, nous faisons de même avec des données parmi lesquelles on ne trouve pas d'effet, situation fréquemment observée dans les programmes d'essais. Les comparaisons aident à expliquer les recommandations de méthodes du § 4.3.

Nous pourrions faire appel aux exemples des tableaux 2 et 3 pour évaluer d'autres programmes statistiques qui commencent à être offerts aux expérimentateurs.

4.4.1 Estimations faites à l'aide de « bonnes » données

Les ensembles hypothétiques de données présentés dans le tableau 2 et illustrés dans la fig. 8 peuvent être qualifiés de « bons » parce qu'ils renferment au moins deux effets partiels, ce qui fait qu'ils sont analysables par régression logit ou probit. L'exemple A illustre plusieurs méthodes d'essai publiées par Environnement Canada. Les trois premiers exemples, de A à C, comportent des données plutôt régulières, tandis que l'exemple D est erratique.

La plupart des programmes statistiques actuels présentent des estimations semblables de la CE 50 dans le tableau 2, particulièrement dans le cas des données régulières. Ces estimations obtenues par l'ordinateur concordent également avec les estimations graphiques fondées sur le bon sens et présentées dans la première ligne du tableau. La fig. 8 montre que les estimations graphiques et informatiques sont raisonnables.

À l'égard des cinq programmes informatiques d'estimation par la méthode des probits, le tableau 2 révèle que les CE 50 sont identiques pour les exemples A, B et C, qui s'appuient sur des données assez régulières. Les limites de confiance étaient également très semblables. Le programme SAS employant la méthode du maximum de vraisemblance pourrait être considéré comme donnant la meilleure évaluation et comme étant la norme de comparaison. Même avec les données irrégulières de l'exemple D, les CE 50 estimées à l'aide des cinq programmes sont assez rapprochées les unes des autres. Les estimations obtenues par les programmes de Stephan, de l'OMEE et CETIS correspondent très étroitement aux estimations du programme SAS. Les programmes TOXSTAT et TOXCALC ont donné une limite supérieure de confiance qui était considérablement plus basse que celle des autres méthodes, pour les données de l'exemple D.

Sebaugh (1998) a effectué une comparaison plus approfondie des programmes de régression probit à l'aide de 50 ensembles de données. Elle a adopté la méthode SAS comme norme et elle a constaté que les CE 50 différaient de plus de 1 % dans 3 cas avec le programme TOXCALC, dans 5 cas avec le programme TOXSTAT et dans 7 cas avec le programme de Stephan. La plupart des comparaisons étaient proches de manière satisfaisante. Un programme de régression probit largement distribué en tant que « gratuiciel » a été compilé par l'USEPA (description dans USEPA, 1995, annexe H) et il a concordé avec le programme SAS sur les 50 ensembles de données.

La méthode de Spearman-Kärber (dite la S.-K.) peut parfois donner des réponses qui sont en accord étroit avec celles de la régression probit. Pour les « bonnes » données des exemples A et B, l'accord avec la régression probit est maintenu, que les estimations par la méthode S.-K. soient obtenues par le programme du MEEO (OMEE) ou le programme TOXSTAT et qu'il y ait 10 % d'équeutage ou qu'il n'y en ait pas (tableau 2).

La S.-K. sans équeutage a donné des résultats insatisfaisants dans l'exemple C du tableau 2, même si cet exemple était constitué de données régulières. Le problème provient de l'absence d'effets à 0 % et à 100 %. Sans eux, les deux programmes ont donné, pour la S.-K. sans équeutage, des estimations aberrantes de la CE 50, et le programme du MEEO (OMEE) n'a pas donné de limites de confiance. Toujours dans l'exemple C, après équeutage de 20 % de la distribution des données, le programme TOXSTAT a estimé la CE 50 à 13,4, près de la valeur « correcte » de 12,6. Le programme du MEEO (OMEE) est aussi arrivé à la même valeur, avec équeutage de 10, 20, 30 ou 35 % (non montré dans le tableau 2). L'estimation de ce paramètre de toxicité convient assez bien. Il semble ainsi que l'équeutage peut être utile à l'estimation convenable du paramètre avec la méthode de S.-K.. On dit généralement que, pour cette méthode, les effets à 0

et à 100 % sont « indispensables ». Cet exemple montre que le programme fonctionne sans ces valeurs, mais ne donne une estimation convenable du paramètre que lorsque l'équeutage permet à d'autres valeurs extrêmes (ici, à 10 % et à 90 % d'effet) de remplacer les effets à 0 et à 100 %.

Des exemples encore plus extrêmes, auxquels manquent les effets nul et total, peuvent être ajustés de manière satisfaisante par la S.-K. avec équeutage. Par exemple, on a postulé un ensemble de résultats qui ne correspondaient qu'à trois concentrations, et les effets étaient de 20, de 50 et de 80 %. La S.-K. sans équeutage a donné une CE 50 absurdement faible, mais après équeutage de 20 %, elle a estimé un paramètre de toxicité et des limites de confiance convenables (TOXSTAT, équeutage automatique minimal, non montré dans le tableau 2). Cet exemple extrême montre aussi que l'équeutage peut être une opération utile avec la S.-K.

Apparemment, la S.-K. sans équeutage peut aussi échouer ou donner des réponses curieuses pour des données modérément ou fortement erratiques. Dans l'exemple D du tableau 2, les méthodes sans équeutage ont donné des estimations grossièrement divergentes de la CE 50 (4,29 et 5,05 au lieu de 26,2, dans la dernière colonne du tableau 2). Manifestement, cela diffère non seulement de la réponse « correcte » donnée par le logiciel SAS, mais cela s'écarte aussi de l'estimation pleine de bon sens, obtenue par une méthode graphique manuelle. De fait, les CE 50 estimées par la S.-K. sans équeutage étaient inférieures à la concentration expérimentale la plus faible, n'ayant causé qu'un effet observé de 10 %. Les programmes du MEEO (OMEE) et TOXSTAT appliquant la S.-K. sans équeutage n'ont pas donné de bons résultats avec l'exemple D (ainsi qu'avec l'exemple C), faute, vraisemblablement, d'effets de 0 et de 100 %.

L'équeutage de 35 % des données irrégulières à chaque extrémité de la distribution, dans l'exemple D, a donné une estimation convenable de 24 (par TOXSTAT, par rapport à la valeur « correcte » de 26,2). Ici encore, l'équeutage a compensé partiellement l'absence d'effets nul et total. Le programme du MEEO (OMEE) a continué de donner des réponses aberrantes, quel qu'ait été l'équeutage entre 10 et 35 % (non montré dans le tableau 2).

D'après ces exemples, il semble qu'avec le programme de S.-K. les estimations devraient être faites avec et sans équeutage. Pour ce qui concerne le taux d'équeutage, l'expérimentateur devrait choisir l'option diversement appelée *« automatic trim », « minimal trim »* ou *« automatically minimize trim level »* (c'est-à-dire *équeutage automatique minimal*) dans les programmes informatiques commerciaux (TOXSTAT, CETIS). Les programmes choisissent le taux approprié. On devrait évaluer les résultats par inspection des données brutes et des représentations graphiques de ces données, puis, après comparaison, choisir les estimations les plus convenables qu'aura données la S.-K., avec et sans équeutage. Ce jugement subjectif de l'expérimentateur n'est pas idéal, mais il semble nécessaire pour les programmes de S.-K., qui n'englobent aucun test de validité du paramètre estimé de toxicité.

Les irrégularités de la méthode de S.-K. n'ont pas d'effet crucial sur les « bonnes » données du tableau 2, parce que la S.-K. ne serait pas utilisée avec ces données, en vertu des méthodes publiées par Environnement Canada. Les quatre exemples seraient normalement analysés par la méthode des logits ou celle des probits. L'exercice précédent visait à évaluer les méthodes de S.-K.

Dans le tableau 2, on s'est également servi de la méthode binomiale, uniquement pour éclairer le propos, parce que tous ces exemples pouvaient être analysés par la régression probit ou logit. Les estimations obtenues par la méthode binomiale excédaient de 6 à 11 % celles de la méthode des probits du logiciel SAS, pour les exemples A, B et C. Évidemment, les limites approximatives de l'estimation diffèrent sensiblement des limites de confiance de la méthode des probits. Avec les données irrégulières de l'exemple D, l'approximation par la méthode binomiale a échoué. Le programme a simplement abouti à la déclaration selon laquelle la CE 50 serait supérieure à la plus faible concentration expérimentale. Il a formulé l'avertissement suivant (traduction) : « L'obtention d'une CL 50 approximative par interpolation entre deux concentrations ne semble pas convenable avec cette [*sic*] donnée ».

Les analyses fondées sur les transformations de Gompertz et arc sinus sont montrées dans le tableau 2,

Tableau 2. — Quatre exemples d'ensembles de données quantiques pour des essais de toxicité aiguë. Voir le texte pour l'explication des méthodes utilisées pour l'analyse.

		Concentration (poids/litre)	Nombre d'organismes touchés (par ex. morts) sur 10			
			Exemple A	Exemple B	Exemple C	Exemple D
		56	-	10	-	5
		32	-	7	9	8
		18	10	5	8	3
		10	9	4	2	4
		5,6	4	0	1	1
		3,2	2	-	1	--
		1,8	0	-	-	--
		Témoin	0	0	0	0
Estimation par la méthode graphique	CE 50		5,6	17	13	29
Probit, maximum de vraisemblance (logiciel SAS)	CE 50 (limites de conf.)		5,58 (4,26-7,40)	16,9 (11,8-23,7)	12,6 (9,02-18,7)	26,2 (13,1-179)
Probit (Stephan et OMEE). La valeur calculée de χ^2 est suivie de sa valeur critique pour $p = 0,05$ et 3 degrés de liberté.	CE 50 (limites de conf.)		5,58 (4,24-7,37)	16,9 (11,9-23,7)	12,6 (8,98-18,6)	26,6 (13,2-187)
	Pente (de la droite)		4,71	3,17	3,07	1,32
	χ^2 (valeur critique)		1,11 (7,82)	3,56 (7,82)	3,47 (7,82)	5,52 (7,82)
Probit (CETIS 1,018)	CE 50 (limites de conf.)		5,58 (4,24-7,37)	16,9 (11,9-23,7)	12,6 (8,98-18,5)	26,6 (13,2-190)
Probit (TOXSTAT 3.5)	CE 50 (limites de conf.)		5,58 (4,38-7,12)	16,9 (12,4-22,9)	12,6 (9,13-17,4)	26,6 (13,4-53,0)
Probit (TOXCALC 5.0)	CE 50 (limites de conf.)		5,58 (4,24-7,37)	16,9 (11,9-23,7)	12,6 (8,98-18,5)	27,6 (15,9-85,7)
Logit (TOXSTAT 3.5)	CE 50 (limites de conf.)		5,63 (4,39-7,22)	16,8 (12,1-23,3)	12,8 (9,36-17,6)	26,5 (13,3-53,1)
Spearman-Kärber, sans équeutage (OMEE)	CE 50 (limites de conf.)		5,64 (4,38-7,26)	16,8 (12,4-22,9)	7,98 (non estimées)	4,29 (non estimées)
Spearman-Kärber, sans équeutage (TOXSTAT 3.5)	CE 50 (limites de conf.)		5,64 (4,40-7,23)	16,8 (12,5-22,7)	10,1 (4,8-21,0)	5,05 (1,39-18,3)
Spearman-Kärber, 10 à 35 % d'équeutage (TOXSTAT 3.5)	CE 50 [% d'équeut.] (limites de conf.)		5,73 [10 %] (2,55-12,9)	16,7 [10 %] (8,30-33,5)	13,4 [20 %] (11,3-15,9)	24,0 [35 %] (16,1-35,8)
Méthode binomiale (Stephan)	CE 50 interpolée (fourchette)		6,22 (1,8-10)	18 (5,6-56)	13,4 (5,6-32)	> 5,6 (avec avertissement)
Gompertz (Weibull) [CETIS 1.018]	CE 50 (limites de conf.)		6,11 (4,43-7,80)	18,6 (12,0-25,2)	14,1 (9,58-19,0)	28,6 (11,2-235)
Transformation arc sinus (CETIS 1.018)	CE 50 (limites de conf.)		5,54 (4,42-7,47)	17,0 (12,8-22,2)	12,1 (8,81-17,7)	26,8 (14,1-153)
Moyenne mobile (Stephan/OMEE)	CE 50 (limites de conf.)		5,58 (4,24-7,33)	17,2 (12,9-22,4)	13,4 (9,0-24,2)	17,8 (11,9-37,1)

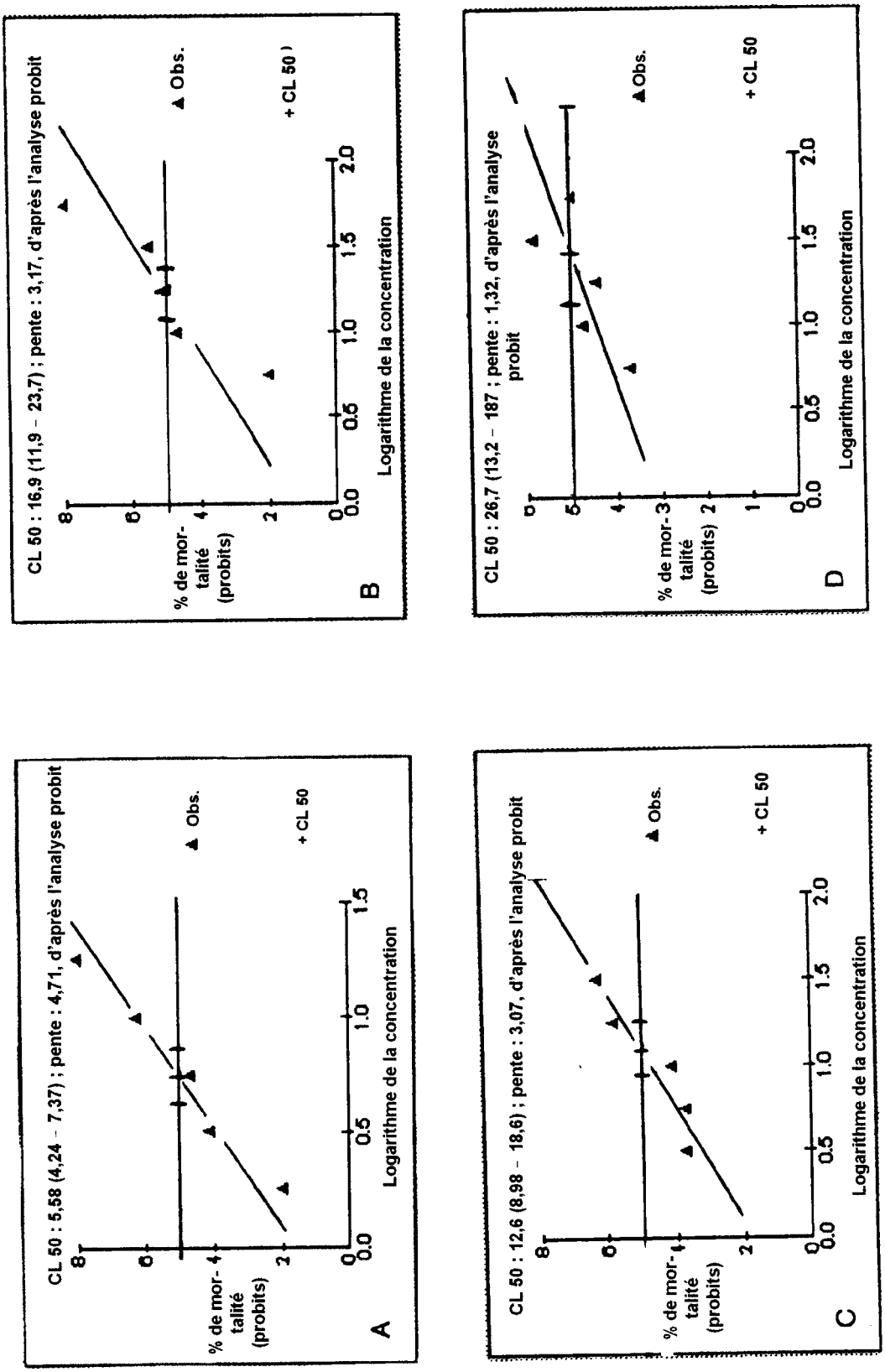

CL 50 : 5,58 (4,24 – 7,37) ; pente : 4,71, d'après l'analyse probit
% de mor-talité (probits)
Obs.
+ CL 50
Logarithme de la concentration
A
CL 50 : 16,9 (11,9 – 23,7) ; pente : 3,17, d'après l'analyse probit
% de mor-talité (probits)
Obs.
+ CL 50
Logarithme de la concentration
B
CL 50 : 12,6 (8,98 – 18,6) ; pente : 3,07, d'après l'analyse probit
% de mor-talité (probits)
Obs.
+ CL 50
Logarithme de la concentration
C
CL 50 : 26,7 (13,2 – 187 ; pente : 1,32, d'après l'analyse probit
% de mor-talité (probits)
Obs.
+ CL 50
Logarithme de la concentration
D

Figure 8. — Aspect graphique des régressions probit correspondant aux exemples A à D du tableau 2. Les graphiques ont été imprimés par le programme informatique du ministère de l'Environnement et de l'Énergie de l'Ontario (MEEO) [OMEE, 1995], et on y a ajouté les éléments suivants : une ligne horizontale au probit 5, les limites de confiance au seuil de 95 % et une droite ajustée des probits.

bien que ces méthodes soient rarement utilisées. Les CE 50 estimées par la méthode de Gompertz sont notablement plus élevées que celles que l'on estime par d'autres méthodes et plus élevées que l'estimation graphique, pleine de bon sens, des exemples A, B et C. Le modèle de Gompertz convient davantage que la transformation normale et logistique, si la répartition des effets est asymétrique. L'analyse de Gompertz est analogue à l'emploi du modèle de Weibull, qui, parfois, se révèle donner le meilleur ajustement aux données de survie (Newman, 1995, p. 125). Le modèle de Weibull suppose aussi une distribution asymétrique. Christensen (1984) a constaté qu'une transformation de Weibull fournissait généralement un ajustement au moins aussi bon aux données expérimentales que le modèle probit. Cependant cela n'est pas évident d'après les CE 50 du tableau 2.

La transformation arc sinus (parfois appelée *transformation angulaire*) a donné des estimations très semblables aux résultats du logiciel SAS et d'autres méthodes des probits. D'après ce fait, la transformation arc sinus semblerait valable, mais elle ne serait pas nécessaire si on disposait d'une bonne méthode des probits ou des logits.

Les programmes employant la méthode de la moyenne mobile de Stephan *et al.* (1978) et du MEEO (OMEE, 1995) ont donné des estimations identiques, qui étaient également identiques ou presque aux estimations par la méthode des probits, avec les « bonnes » données des exemples A à C. Cependant, la méthode de la moyenne mobile a donné une CE 50 et des limites de confiance plutôt aberrantes pour les données irrégulières de l'exemple D. Comme nous l'avons déjà dit, la méthode ne semblerait pas nécessaire dans les circonstances normales, parce que le programme disponible a les mêmes exigences concernant le type de données que la méthode des probits et celle des logits.

4.4.2 Estimations avec des données comportant peu d'effets partiels

Le plus souvent, les laboratoires obtiennent des résultats expérimentaux ne correspondant à aucun ou à un seul effet partiel. Les résultats ne peuvent pas être analysés par régression probit ou logit. L'utilité d'autres méthodes est évaluée grâce aux exemples du tableau 3.

Les données du tableau 3 ont été obtenues à partir de celles du tableau 2, par réduction de la plupart des exemples à un effet partiel. Les deux valeurs aux concentrations maximales ont été fixées à 100 % d'effet, tandis que les deux valeurs aux concentrations minimales basses ont été fixées à 0 % d'effet. La seule exception se trouve dans l'exemple D, où on a maintenu la valeur irrégulière à 50 % d'effet à la forte concentration. Les méthodes énumérées dans la colonne de gauche ont servi à analyser ou à tenter d'analyser ces données. Conformément à la recommandation formulée dans le § 4.2, les analyses n'ont utilisé qu'un des deux effets successifs de 0 ou de 100 %, celui qui était le plus près du centre.

Les exemples A, B et C ne peuvent pas être analysés par la régression probit ou logit. La méthode de la moyenne mobile ne peut pas non plus donner de réponse, confirmant qu'elle n'est pas d'une grande aide aux expérimentateurs en tant que méthode de secours.

Les analyses par les méthodes des probits et des logits ont fonctionné de manière satisfaisante avec les données irrégulières de l'exemple D. Les cinq programmes d'estimation par la méthode des probits ont donné la même CE 50 convenable, et l'estimation par la méthode des logits suivait de près. Les limites de confiance variaient quelque peu ; celles des méthodes de Stephan et du MEEO (OMEE) allaient de 0 à l'infini, ce qui n'est pas très utile.

Pour les exemples A, B et C, la méthode binomiale et celle de Spearman-Kärber sans équeutage ont donné des estimations qui semblaient convenables et qui concordaient assez bien avec les estimations obtenues par une méthode graphique manuelle. Cela conforte la pratique récemment instituée par Environnement Canada d'utiliser la S.-K. quand il n'y a qu'un seul effet partiel, ce qui empêche l'emploi des probits ou des logits (EC, 2001a ; 2004a). À noter que les analyses réussies grâce à la S.-K. concernaient des données renfermant des effets à 0 et à 100 %. La S.-K. avec équeutage a failli dans chacun de ces trois exemples (TOXSTAT) ou a donné des estimations quelque peu divergentes (OMEE), vraisemblablement parce que l'équeutage n'était pas approprié au petit nombre d'observations.

Dans l'exemple D, la méthode de Spearman-Kärber et la méthode binomiale ne seraient pas nécessaires, puisque les méthodes privilégiées de régression probit ou logit ont donné des estimations de la CE 50 et de ses limites de confiance. Cependant, il vaut la peine d'examiner les performances de ces méthodes de second rang. Ni la S.-K. sans équeutage ni la méthode binomiale n'a pu fonctionner avec les données irrégulières de l'exemple D. La S.-K. sans équeutage a abouti à une CE 50 qui était exagérément faible par rapport aux valeurs données par la régression probit ; les programmes TOXSTAT et du MEEO (OMEE) ont donné la même CE 50 absurde. L'estimation de la CE 50 par TOXSTAT, après équeutage, était du bon ordre de grandeur, mais un peu basse. Les résultats erratiques de la méthode de S.-K. à l'égard de l'exemple D confortent la recommandation récente d'Environnement Canada selon laquelle il ne faut l'utiliser que lorsque la régression probit ou logit n'a pas fonctionné, en raison d'un seul effet partiel.

Les résultats obtenus par la S.-K. avec les données de l'exemple D montrent également que l'analyse avec ou sans équeutage est nécessaire et que l'on devrait choisir entre les deux après avoir comparé les résultats bruts. Dans certains cas, les estimations des deux variantes de la méthode pourraient ne pas être convenables, et l'expérimentateur pourrait devoir user de son jugement et les rejeter. Dans les logiciels disponibles, il ne semble pas y avoir de règle fixe, applicable à détermination du caractère acceptable des résultats de la S.-K., ni de test de validité. Il faut donc continuer de faire appel à son jugement.

Pour ce qui concerne la méthode de S.-K. du programme du MEEO (OMEE), les taux d'équeutage supérieurs à 10 % ont donné des estimations de la CE 50 de plus en plus hautes et de moins en moins convenables avec les données des exemples A, B et D ainsi que des résultats erratiques avec les données de l'exemple C (non montré dans le tableau 3). La méthode de S.-K. du programme du MEEO (OMEE) semble avoir un défaut, et nous recommandons à l'expérimentateur d'utiliser les versions disponibles de la méthode dans les progiciels commerciaux.

4.5 Examen des méthodes statistiques d'estimation des CE p

Repères

- *Les effets quantiques suivent une loi binomiale, et l'analyse doit utiliser des méthodes appropriées. On utilise couramment, par tradition, la transformation en probits de l'effet quantique pour linéariser la relation avec le logarithme de la concentration. Mathématiquement, la transformation en logits est supérieure et elle donne des estimations similaires, bien que, par le passé, elle ait été moins souvent utilisée par les écotoxicologues.*
- *Dans la régression probit ou logit, les estimations du maximum de vraisemblance sont définitives et elles ont comme grande qualité de séparer tout effet observé chez les témoins de manière avantageuse. Cependant, les méthodes du maximum de vraisemblance sont généralement offertes uniquement dans les gros progiciels, ce qui fait qu'elles ne sont pas souvent utilisées pour les analyses en routine en écotoxicologie.*
- *La régression probit ou logit classique procède par une succession d'ajustements améliorés d'une droite (itération). Le test du khi-deux permet de juger du caractère acceptable de l'ajustement.*
- *Dans un nombre limité d'essais, la transformation arc sinus s'est également révélée satisfaisante.*
- *Les méthodes de raccourci graphique de Litchfield-Wilcoxon de régression probit sont périmées, mais elles pourraient être utiles pour vérifier les résultats du traitement informatique ou pour former le nouveau personnel.*
- *La méthode de Spearman-Kärber (la S.-K.) ne permet pas d'estimer les paramètres de toxicité par régression, mais par les moyennes pondérées des points intermédiaires entre les concentrations logarithmiques. Elle exige des données symétriques monotones et des effets de 0 et de 100 %. Si les données ne sont pas monotones, les programmes d'analyse peuvent imposer un lissage. Si on ne possède pas de données sur les effets nul et total, l'équeutage des données pourrait aboutir à des estimations satisfaisantes*

Tableau 3. — Quatre exemples d'ensembles de données quantiques avec quelques effets partiels. Voir le texte pour l'explication des méthodes utilisées pour l'analyse.

		Concentration (poids/litre)	Nombre d'organismes touchés (par ex. morts) sur 10			
			Exemple A	Exemple B	Exemple C	Exemple D
		56	-	10	-	5
		32	-	10	10	10
		18	10	5	10	3
		10	10	0	2	0
		5,6	4	0	0	0
		3,2	0	-	0	--
		1,8	0	-	-	--
		Témoin	0	0	0	0
Estimation par la méthode graphique	CE 50		6,1	18	12,4	31
Probit (SPSS)	CE 50 (limites de conf.)		-----	-----	-----	28,4 (17,9-28,2) *
Probit (Stephan et OMEE).	CE 50 (limites de conf.)		-----	-----	-----	28,4 * (0- ∞)
Probit (CETIS 1,018)	CE 50 (limites de conf.)		-----	-----	-----	28,4 (non estimées)
Probit (TOXSTAT 3.5)	CE 50 (limites de conf.)		-----	-----	-----	28,4 * (19,4-41,5)
Probit (TOXCALC 5.0)	CE 50 (limites de conf.)		-----	-----	-----	28,4 * (non estimées)
Logit (TOXSTAT 3.5)	CE 50 (limites de conf.)		-----	-----	-----	27,6 (18,7-40,8)
Spearman-Kärber, sans équeutage (OMEE, TOXSTAT)	CE 50 (limites de conf.)		5,96 (4,99-7,11)	17,9 (14,9-21,6)	11,9 (10,3-13,8)	9,11 (5,25-25,5)
Spearman-Kärber, 10 à 35 % d'équeutage (TOXSTAT 3.5)	CE 50 (limites de conf.)		-----	-----	-----	23,2 [30 %] (18,1-29,9)
Spearman-Kärber, 10 d'équeutage (OMEE)	CE 50 (limites de conf.)		7,02 (5,61-8,79)	24,1 (19,1-30,4)	15.5 (12,6-19,1)	15,8 (---------)
Méthode binomiale (Stephan)	CE 50 interpolée (fourchette)		6,03 (3,2-10)	18 (10-32)	12,0 (5,6-18)	> 5,6 (avec avertissement)
Moyenne mobile (Stephan/OMEE)	CE 50 (limites de conf.)		-----	-----	-----	17,8

* Dans le cas de l'exemple D, tous les programmes fondés sur la méthode des probits et des logits ont lancé un avertissement d'hétérogénéité significative ; la plupart ont mis en doute la validité des limites de confiance. Le programme fondé sur la méthode de la moyenne mobile a lancé un avertissement selon lequel les limites de confiance étaient probablement trop rapprochées. Celui de la méthode binomiale a lancé un avertissement selon lequel l'interpolation ne semblait pas « raisonnable ».

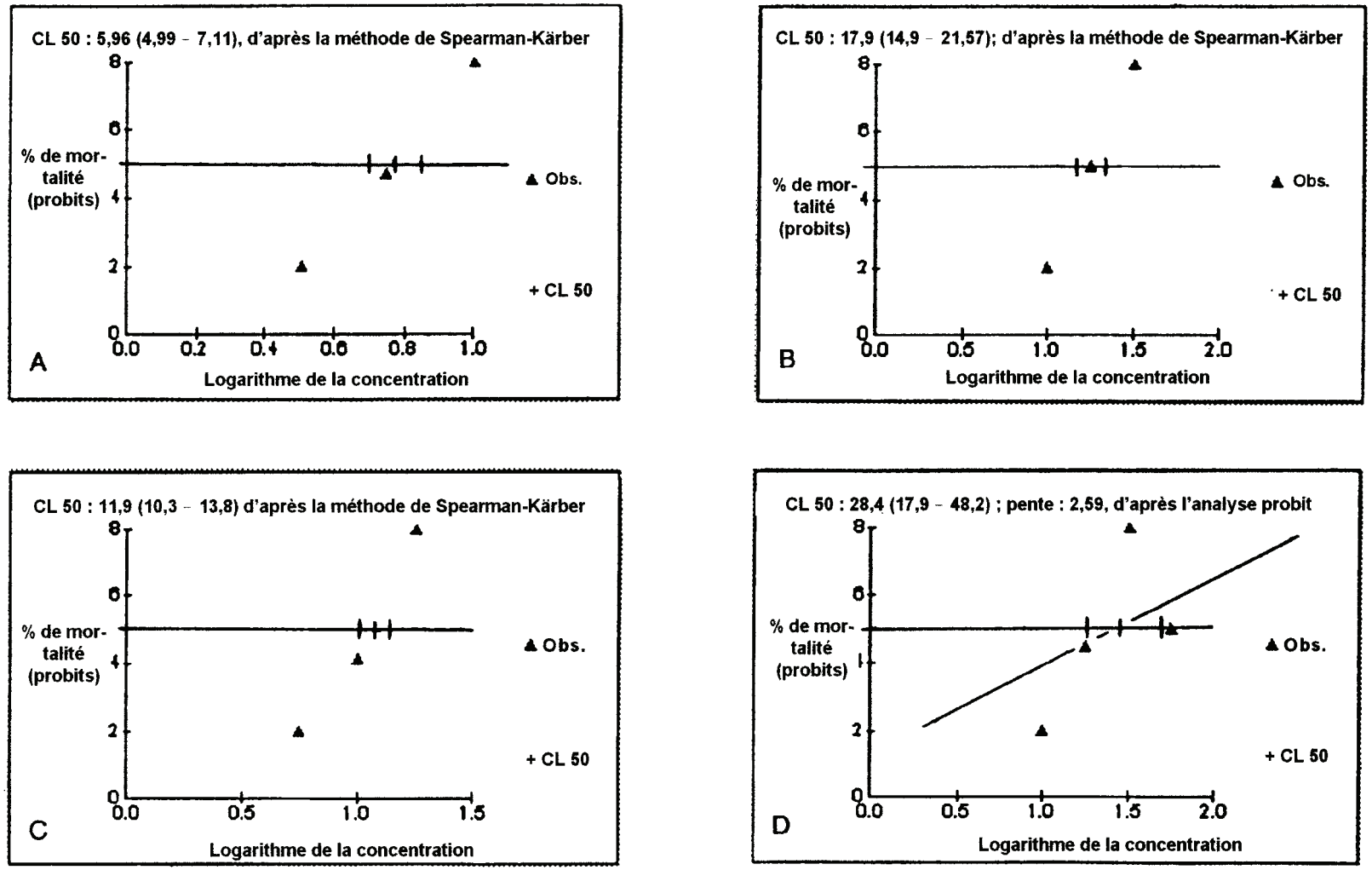

CL 50 : 5,96 (4,99 - 7,11), d'après la méthode de Spearman-Kärber
% de mortalité (probits)
Obs.
+ CL 50
A
Logarithme de la concentration
CL 50 : 17,9 (14,9 - 21,57); d'après la méthode de Spearman-Kärber
% de mortalité (probits)
Obs.
+ CL 50
B
Logarithme de la concentration
CL 50 : 11,9 (10,3 - 13,8) d'après la méthode de Spearman-Kärber
% de mortalité (probits)
Obs.
+ CL 50
C
Logarithme de la concentration
CL 50 : 28,4 (17,9 - 48,2) ; pente : 2,59, d'après l'analyse probit
% de mortalité (probits)
Obs.
+ CL 50
D
Logarithme de la concentration

Figure 9.— Graphiques de données quantiques comportant quelques effets partiels (tableau 3). Les graphiques ont été imprimés par le programme informatique du ministère de l'Environnement et de l'Énergie de l'Ontario (MEEO) [OMEE,1995]. Pour les graphiques A à C, les estimations ont été obtenues par la méthode de Spearman-Kärber sans équeutage. On ne voit pas de droites ajustées parce que cette méthode n'en utilise pas. L'analyse des données du graphique D a été effectuée par le programme de régression probit du logiciel SPSS.

à partir de certains ensembles de résultats. Certaines méthodes d'essai publiées récemment par Environnement Canada préconisent l'emploi limité de seulement la S.-K. sans équeutage. Il semble souhaitable d'effectuer l'analyse avec un équeutage minimal ou sans équeutage, puis de juger de l'acceptabilité de chaque paramètre de toxicité ainsi estimé par comparaison avec les données brutes.

- *Pour ce qui concerne les essais auxquels ne correspondent pas d'effets partiels, la méthode binomiale permet d'estimer une CE 50 approximative en tant que moyenne géométrique des concentrations ne causant aucun effet et causant un effet total et de prendre ces concentrations comme limites, à l'intérieur desquelles se trouvent les limites de confiance.*
- *La méthode de la moyenne mobile fonctionne généralement bien, mais elle est superflue, parce que le programme informatique disponible exige deux effets partiels et que, à la place, on peut utiliser la régression probit ou logit.*
- *Aux États-Unis, l'« interpolation linéaire » a été désignée technique particulière. Elle équivaut essentiellement à la méthode binomiale. Les expérimentateurs devraient se méfier de certains vieux programmes informatiques employant cette méthode, qui n'utilisent pas les logarithmes de la concentration.*
- *On fournit une liste de critères pour évaluer les nouveaux programmes informatiques d'analyse des données quantiques.*
- *À l'avenir, les analyses pourraient utiliser la régression non linéaire si des progiciels commodes sont offerts aux écotoxicologues.*

4.5.1 Régressions probit et logit en général

La régression probit ou logit est une méthode couramment utilisée et satisfaisante d'analyse des données quantiques. Mathématiquement, les logits sont supérieurs, comme nous l'expliquons dans l'annexe J, mais les probits ont été couramment utilisés en écotoxicologie. À l'instar de toutes les autres méthodes, celle-ci est des plus efficaces si les données sont assez lisses et régulières et elle exige deux effets partiels. La droite log-probit ajustée à vue d'œil (§ 4.2.2 et fig. 5) est une forme de régression probit, effectuée mentalement, sans calculs.

Il faut expliquer pourquoi on passe d'une distribution binomiale (pour les données quantiques) à une analyse fondée sur une distribution normale (comme dans la régression probit).

1. Pour les données quantiques telles que les résultats des essais de toxicité létale, la mortalité d'un organisme est un phénomène binaire, en tout ou rien (oui ou non).

2. Dans un récipient, le nombre d'organismes touchés (y) est la somme des résultats binaires individuels. La variable y est une variable aléatoire binomiale. Pour ce récipient, les résultats expérimentaux s'expriment par y (le nombre d'organismes touchés) divisé par n (le nombre d'organismes dans le récipient).

3. Habituellement, plusieurs récipients sont exposés à différentes concentrations. Si les proportions d'organismes touchés dans chaque récipient sont portées sur un graphique, en fonction du logarithme de la concentration, et si on réunit les points, la relation dose-effet empirique qui s'en dégage ressemble à une fonction de répartition de la loi normale (fig. 10, graphique de gauche). Elle ressemble aussi à une courbe cumulative de la distribution logit (fig. 10, graphique de droite) ou à une distribution de Gompertz. Cette distribution décrit la résistance de l'échantillon d'organismes au toxique.

4. Cette distribution peut désormais être traitée comme normale ou logistique, etc. Les effets binomiaux dans la distribution sont transformés à l'aide des transformations probit, logit ou de Gompertz, etc., qui redressent la courbe dose-effet sigmoïde (fig. 10).

5. La relation linéaire résultante entre le logarithme de la concentration et l'effet binomial sert à estimer les ordonnées à l'origine et les pentes. Ensuite, on utilise le modèle linéaire à la manière d'une régression inverse (v. le § 9.4) pour estimer la CE p.

La régression logistique et probit sont deux méthodes courantes que l'on emploie pour la transformation de l'étape 4 ; les transformations en question sont montrées dans la fig. 10 et décrites plus en détail dans les annexes H et J. Les formules mathématiques des modèles probit, logit et de Weibull sont présentées et expliquées dans OECD (2004).

La partie gauche de la fig. 10 montre schématiquement le calcul des probits. La courbe est une courbe typique du pourcentage d'effet en fonction du logarithme de la concentration. Les tiretés horizontaux représentent les écarts types de la courbe normale cumulative (écarts types par pas de 0,5 sur l'échelle verticale du pourcentage d'effet). À partir de leur point d'intersection avec la courbe, on fait descendre des lignes verticales sur une échelle uniformisée des écarts types. Les unités de cette échelle s'appellent *variables normales équivalentes* (ou *normits* ou *NED,* pour *normal equivalent deviates*). Sur l'échelle, 0 NED correspond à l'effet de 50 % ; au-dessus et au-dessous, les valeurs sont respectivement positives et négatives, comme on peut le lire au bas des lignes verticales. Pour faciliter le traitement mathématique à l'époque où les calculs se faisaient à la main, on a ajouté aux NED la valeur de 5 et on a appelé le résultat *probits,* figurant dans le bas de la partie gauche de la figure. Si on espace régulièrement les probits sur l'axe vertical, on redresse la courbe, qui devient une droite en fonction du logarithme de la concentration (montrée dans l'annexe H).

Les logits sont montrés dans la partie droite de la fig. 10. Les mêmes explications valent, sauf que la répartition des effets est posée comme étant logistique plutôt que normale. Les tiretés horizontaux expriment des logits. Prolongés verticalement à partir du point d'intersection avec la courbe, ils aboutissent sur l'échelle horizontale des logits dans le bas de la figure. Le résultat est similaire ; la courbe est redressée lorsque l'on exprime les logits en fonction du logarithme de la concentration.

Après la transformation en probits (ou en logits), on procède à l'analyse statistique. Comme on le décrit dans le texte qui suit, il faut estimer les paramètres du modèle probit ou logit par des procédés plutôt complexes, et, à cette fin, l'emploi de programmes informatiques est universel.

4.5.2 Autres transformations

Il existe d'autres modèles et transformations. Dans les calculs, on peut employer la transformation arc sinus et celle de Gompertz, qui sont exactement analogues aux transformations utilisées avec les probits. Ces méthodes sont frappées des mêmes limites que celles qui s'appliquent aux probits, notamment la nécessité de connaître deux effets partiels.

Dans le tableau 2, on a montré les résultats de l'analyse faite avec la transformation de Gompertz, et les CE 50 étaient notablement plus fortes que celles que l'on avait estimé par d'autres méthodes et elles étaient généralement plus fortes que les estimations pleines de bon sens obtenues par la méthode graphique. Comme nous l'avons mentionné dans le § 4.4.1, le modèle de Gompertz et le modèle analogue de Weibull conviendraient davantage à des données dans lesquelles les effets seraient asymétriques. La transformation arc sinus a permis d'estimer des CE 50 semblables à celles des méthodes des probits. On pourrait utiliser cette transformation, mais celle-ci ne semblerait pas nécessaire si on pouvait disposer d'une bonne méthode des probits ou des logits.

4.5.3 Régression probit classique informatisée

Dans les programmes informatiques actuels de régression probit, l'opérateur saisit les données brutes (arithmétiques), et, généralement, les programmes font les transformations appropriées en logarithmes de la concentration et en probits de l'effet. Certains programmes ont leur « tempérament ». Avec TOXSTAT 3.5, l'opérateur doit spécifier la transformation en logarithme de la concentration et, immédiatement après, il doit commander l'exécution de la transformation, sinon celle-ci sera oubliée pendant les calculs. TOXCALC a le défaut de ne pas utiliser les probits dans le graphique qu'il trace.

La plupart des programmes informatiques actuels de régression probit suivent les méthodes « classiques » créées pour les calculatrices mécaniques d'avant l'avènement des ordinateurs (Finney 1971 ; Hubert 1992). La transformation de l'effet proportionnel en probits signifie que la relation est linéarisée en fonction du logarithme de la dose et que l'ajustement est simplifié en une régression linéaire pondérée. L'ajustement est obtenu par approximations successives jusqu'à l'obtention de la meilleure droite

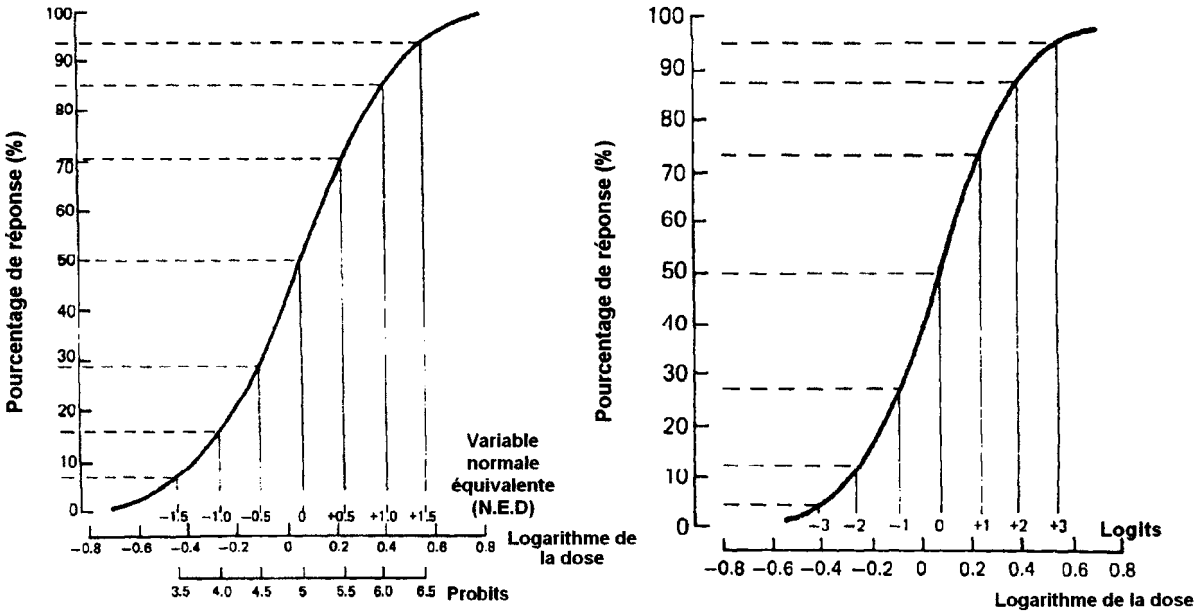


Figure 10. — Démonstration graphique des transformations en probits et en logits (d'après Hewlett et Plackett, 1979)

(*itération*), à l'aide d'une technique des moindres carrés.

À l'époque des calculatrices mécaniques, les calculs étaient faisables, mais restaient fastidieux, longs, et ils donnaient prise à l'erreur. Les itérations étaient nécessaires parce que les pondérations (valeurs relatives) des observations étaient initialement inconnues et dépendaient des paramètres qui n'étaient pas encore estimés. Le nombre d'individus à chaque concentration contribue à la pondération. On peut qualifier le processus de « méthode des moindres carrés itérativement repondérés »[32].

Cette méthode classique, lourde si les calculs sont effectués à la main, peut être appliquée sans effort par les ordinateurs. Ce n'est pas une estimation du maximum de vraisemblance, mais les résultats sont essentiellement les mêmes ; c'est-à-dire que l'on « parvient à une solution qui possède un maximum de vraisemblance » pour l'estimation de la CE 50 et les

32. Ces opérations mathématiques, conçues en raison des capacités limitées des calculatrices mécaniques, sont assez complexes. (1) L'ordinateur ajuste grossièrement une droite aux données brutes, à l'aide des logarithmes de la concentration et des probits de l'effet. (2) Il « lit » les probits prévus (= % d'effet) sur la droite grossièrement ajustée. (3) Dans un tableau de constantes, il recherche les coefficients initiaux de pondération des observations, d'après les probits prévus, puis affecte ces coefficients aux observations. (4) Il cherche les probits de travail d'après les probits prévus et observés. (5) Il ajuste une meilleure droite, d'après les probits de travail, les coefficients de pondération et le nombre d'organismes. Cela aboutit à la première estimation de la CE 50, des limites de confiance et du khi-deux comme mesure de l'ajustement. (6) Il effectue un autre cycle des étapes 2 à 5, en utilisant les probits de travail (de 4) comme s'ils étaient des données brutes. Autrement dit, il cherche de nouvelles valeurs pour les coefficients de pondération et les probits de travail. (7) le processus 6 se répète tant que les réponses n'approchent pas d'une valeur stable (« convergent »), et les calculs finals sont adoptés.

Cet ajustement par les moindres carrés repondérés et réitérés est une façon de calculer une solution ayant un maximum de vraisemblance. Parfois le nombre de cycles dépend de la volonté de l'opérateur, sinon le programme dispose d'un critère intégré pour interrompre les cycles. Parfois, il suffit de deux ou trois cycles ; pour les bons ensembles de données, les résultats de cycles successifs changent peu. Dans le cas de données irrégulières, il pourrait ne pas y avoir de convergence (absence d'ajustement convenable) au bout de 20 cycles ; il y aurait peu de raisons de poursuivre l'ajustement plus loin. Les données anormales donnent parfois des résultats curieux après une demi-douzaine de cycles, comme une pente très faible, peu réaliste, et un large intervalle de confiance, alors que le programme tente de représenter la gamme des résultats.

paramètres de la droite. La méthode itérative classique a déjà été considérée comme la « plus efficace » à l'égard des bonnes distributions des données qui sont log-normales (Gelber *et al.,* 1985).

La technique itérative des moindres carrés souffre d'un grave handicap. On ne peut pas l'étendre efficacement à tout effet observé chez le témoin. Cela peut être fait par des modèles appropriés, utilisant les techniques du maximum de vraisemblance (§ 4.2.3 et 4.5.5).

4.5.4 Évaluation de l'ajustement avec le khi-deux

L'ajustement de la droite des probits est révélé par la valeur calculée de khi-deux, qui ne doit pas dépasser une valeur critique si l'on veut accepter la droite et les estimations. Les programmes informatiques s'occupent normalement de ces calculs, mais l'expérimentateur devraient s'assurer de la valeur satisfaisante de χ^2. L'évaluation par le khi-deux est approximative, parce qu'elle exigerait au moins 30 individus par traitement « pour être statistiquement justifiée » (Hubert, 1992).

On peut trouver les valeurs critiques de khi-deux dans les manuels classiques de statistique. Dans un essai de toxicité, le nombre de degrés de liberté égale le nombre de concentrations employées moins 2. Le tableau qui suit pourrait servir pour une probabilité de 0,05.

Nombre de degrés de liberté	Valeur critique de χ^2
1	3,54
2	5,99
3	7,82
4	9,49
5	11,1
6	12,6

Dans les quatre exemples du tableau 2, le nombre de concentrations est toujours de cinq. Le nombre de degrés de liberté est donc de trois. La valeur critique de χ^2 est 7,82. Si le χ^2 calculé excédait cette valeur, les données seraient significativement hétérogènes, et la droite ne serait pas un ajustement acceptable. Les quatre exemples du tableau 2 sont tous acceptables.

Il est également souhaitable de vérifier de visu la droite calculée des probits. On devrait la comparer à une droite ajustée à vue d'œil, créée à cette fin (§ 4.2.2). Le programme du MEEO (OMEE) produit un graphique des résultats, et d'autres programmes pourraient faire de même. Sinon, la droite calculée devrait être tracée à côté de la droite tracée à la main, sur papier log-probit. Le tracé peut être facile à réaliser, puisque la pente calculée par le programme représente l'augmentation du nombre de probits pour un cycle logarithmique de concentrations. À partir du point connu de la CE 50, on mesure un cycle logarithmique et un probit (vers le haut, vers le bas ou les deux) pour placer un second point sur la droite (ou les deuxième et troisième points). Le tracé est encore plus facile si un programme informatique produit une liste de paramètres de toxicité (CE 10, CE 20, etc.), comme le font les programmes SAS, SPSS, CETIS, TOXCALC et TOXSTAT.

4.5.5 Estimations du maximum de vraisemblance

La méthode du maximum de vraisemblance (MMV) est une technique objective de sélection de la valeur des paramètres pour un modèle servant à ajuster un ensemble de données. Les paramètres sont choisis pour maximiser (dans un modèle choisi) la probabilité d'observer les données effectivement collectées.

Dans un essai de toxicité quantique, le nombre d'organismes touchés à une concentration donnée suit une distribution binomiale. Les paramètres des distributions binomiales sont censés être reliés aux concentrations par une fonction, d'habitude la normale ou la logistique. Dans ces conditions, les estimations du maximum de vraisemblance se révèlent dépendre de deux équations. Aujourd'hui, on peut résoudre directement les équations pour choisir les valeurs des paramètres, à l'aide d'un ordinateur personnel et de progiciels modernes de statistique tels que SAS [33].

L'utilisation de la méthode du maximum de vraisemblance dans les essais de toxicité ne constitue qu'une petite partie de son application générale. On peut adopter des modèles de divers type pour ajuster différentes données, tandis que les techniques d'estimation du maximum de vraisemblance s'appliquent dans tous les cas. Ainsi, la MMV pourrait servir à l'analyse des résultats de divers types d'essais de toxicité, quantitatifs ou quantiques. Par exemple, le

33. Finney (1978), pionnier du domaine, a salué l'avènement des ordinateurs modernes en déclarant que l'un des gains les plus importants que les ordinateurs avaient procurés à la statistique était de faciliter le lancement et l'exécution de calculs itératifs et que, en outre, les calculs itératifs de régression probit et logit classique pouvaient être remplacés par des techniques d'optimisation directe qui parvenaient aux mêmes réponses rapidement et avec plus de précision qu'auparavant.

modèle pourrait être une régression du poids des organismes sur le logarithme des concentrations d'exposition. Ou, encore, ce pourrait être une fonction décrivant la distribution de probabilité d'un seul ensemble d'observations. Ici, nous envisageons des modèles pour les essais quantiques (c'est-à-dire la régression probit) représentant une application très utile de la MMV.

Pour la régression probit, la MMV est « équivalente » à la vieille méthode des moindres carrés itérativement repondérés (Jennrich et Moore, 1975). Autrement dit, la MMV parvient à estimer une CE 50 et ses limites de confiance qui sont très semblables à celles auxquelles aboutirait la technique itérative décrite dans le § 4.5.3. Toutefois, la MMV est mathématiquement plus élégante et elle devrait être considérée comme la méthode définitive. Dans la vieille régression probit itérative, deux paramètres sont dignes d'intérêt : la pente et l'ordonnée à l'origine. Dans les programmes modernes employant la MMV, ces paramètres sont remplacés par leurs équivalents, la moyenne et la variance. Une fonction de vraisemblance est manipulée de façon à exprimer les paramètres en fonction des données. Par calcul, on fait égaler la dérivée première à zéro, puis on résout les équations pour les estimations du maximum de vraisemblance des paramètres.

Technique usuelle d'analyse statistique, la MMV fait partie des grands progiciels de statistique. La régression probit l'employant est spécifiquement offerte dans le progiciel SAS (2000) et peut-être dans d'autres. Le programme SPSS et les programmes TOXSTAT, TOXCALC et CETIS employés en toxicologie semblent utiliser les vieilles méthodes itératives d'ajustement de droites. Les écotoxicologues trouveraient sans doute commode de faire inclure la MMV dans des progiciels adaptés à leurs besoins.

Effets observés chez les témoins. — Un grand avantage de la régression probit ou logit avec la MMV est leur capacité d'estimer un effet observé chez les témoins comme variable séparée et de n'utiliser que l'effet provoqué par le toxique pour estimer la CE p. Les effets observés sont la somme de deux sources d'effet, et l'amplitude de l'effet observé chez les témoins compte comme l'un des paramètres à déterminer dans le modèle. Le modèle le plus complexe doit résoudre trois équations (celles de la moyenne, de la variance et de l'effet observé chez les témoins). On estime deux intensités d'effet : l'un d'eux est l'effet de base ou l'effet observé chez les témoins, qui n'est pas attribuable au toxique ; l'autre est l'effet progressif du toxique, agissant seul, dont on se sert pour estimer la CE 50 sans l'effet de base, qui se manifeste chez le témoin.

La MMV est la meilleure méthode mathématique pour tenir compte de l'effet observé chez les témoins. Cependant, comme nous le faisons remarquer dans le § 4.2.4, elle ne peut pas remédier à toute interaction qui est biologique plutôt que d'être statistique. Par exemple, la maladie pourrait provoquer un effet chez les témoins et pourrait aussi affaiblir la résistance des organismes en expérience au toxique. L'analyse estimerait une CE p statistiquement valable, mais pour des organismes affaiblis.

4.5.6 Méthode de Spearman-Kärber

Nous recommandons la méthode de Spearman-Kärber (la S.-K.) pour les données quantiques qui englobent : *a*) un effet partiel ; *b*) des effets à 0 et à 100 %. Autrement dit, la méthode peut être utilisée lorsque les méthodes des probits ou des logits ne fonctionnent pas parce que les données ne comprennent pas deux effets partiels. Cette méthode est offerte dans la plupart des programmes commerciaux tels que CETIS et TOXSTAT ainsi que sur le site Web http://www.epa.gov/nerleerd/stat2.htm. Elle est également disponible dans le programme du MEEO (OMEE), bien que cette version semble mal fonctionner dans certains cas où les données sont irrégulières et qu'il soit préférable d'éviter de l'utiliser.

Préconisée pour les essais d'écotoxicité par Hamilton *et al.* (1977), la méthode de Spearman-Kärber procède très différemment, mathématiquement, de la régression probit. Elle estime la CE 50 à partir des moyennes pondérées des *points milieux entre les concentrations,* sur une échelle logarithmique. La pondération appliquée à chaque point milieu est la *modification de la proportion de l'effet* entre les deux concentrations, similairement à l'estimation de la moyenne d'une distribution de fréquences par multiplication des centres de classe par la proportion propre à chaque classe. (Pour de plus amples explications, v. l'annexe K.)

La S.-K. peut fonctionner en dépit de l'espacement inégal des concentrations sur l'échelle logarithmique et, également, avec des nombres inégaux d'organismes à diverses concentrations. Il n'existe pas de méthode intrinsèque pour traiter un effet qui s'est manifesté chez les témoins.

On peut estimer les intervalles de confiance s'il existe au moins un effet partiel. Ils égalent ± 2 écarts types par rapport à la CE 50. Cela suppose que la CE 50 suit une loi normale et se comporte comme une variable aléatoire normale (Miller et Halpern, 1980). Les limites ne sont pas « susceptibles d'être très erronées » à moins que le nombre d'observations ne soit faible (Finney, 1978).

Le test repose sur les exigences ou hypothèses, exposées ci-dessous, de monotonie et de symétrie des données, lesquelles comprennent des effets à 0 et à 100 %.

- **Les données doivent être monotones.** — Si les effets diminuent d'une concentration à une autre, plus forte, alors les effets sont moyennés, et le résultat est attribué aux deux concentrations. Ce lissage est appliqué successivement à l'ensemble de données jusqu'à ce que ces dernières deviennent monotones (annexe K). Le lissage ne modifie pas la valeur calculée de la CE 50, mais il modifie les limites de confiance.

- **La symétrie** est une hypothèse de la méthode. Si la distribution des effets est asymétrique, la S.-K. n'estime pas une vraie CE 50. Même si on employait l'équeutage, l'estimation de la CE 50 ne serait convenable que si la partie centrale (non équeutée) de la distribution était symétrique.

- **Il faut des effets nul et total,** et cela découle quelque peu de l'hypothèse de symétrie. Sans ces effets extrêmes, la méthode sans équeutage échouerait ou, au mieux, donnerait des résultats anormaux. L'équeutage peut parfois remédier à l'absence d'effets nul et total, s'il existe des effets petits et grands, comme à 10 et à 90 %.

- **L'équeutage** est un moyen de tenter de corriger l'absence de symétrie des extrémités de la courbe dose-effet. On peut l'appliquer pour supprimer les valeurs extrêmes et utiliser les données centrales. L'opération peut être utile s'il existe des proportions inopinément grandes d'organismes à l'une ou l'autre des extrémités de la distribution, c'est-à-dire beaucoup d'organismes ayant réagi à la faible concentration ou n'ayant pas réagi à la forte concentration. Hamilton (1979 ; 1980) a étudié ces situations et a constaté qu'un peu d'équeutage réduisait considérablement plus l'erreur type de la CE 50 estimée (c'est-à-dire que la méthode était plus optimiste) que d'autres méthodes de référence telles que l'analyse probit ou logit du maximum de vraisemblance. Un équeutage plus poussé a encore diminué l'erreur type, mais en augmentant l'estimation de la CE 50. Hamilton a proposé un équeutage de 10-20 % dans les cas où des résultats erratiques se trouveraient dans les parties extrêmes de la distribution, mais en évitant l'élaguer les données dont la distribution était régulière.

Les méthodes d'essai publiées récemment par Environnement Canada (EC, 2001a ; 2004a) sont semblables aux recommandations de Hamilton (1979 ; 1980), mais en étant plus restrictives parce qu'elles n'autorisent pas l'équeutage. Au cours d'une réunion du Groupe consultatif sur la statistique d'Environnement Canada, on a exprimé des doutes sur l'ajustement d'un modèle statistique qui feraient paraître les données plus robustes qu'en réalité. On y aurait déclaré qu'on ne pouvait faire d'une buse un épervier et que l'équeutage conduisait à des difficultés avec la variance et, de là avec les limites de confiance » (Miller *et al.,* 1993). Comme nous le mentionnons plus loin, l'interdiction de l'équeutage est probablement dictée par un excès de prudence.

Quoi qu'il en soit, nous recommandons la méthode de Spearman-Kärber, uniquement pour les essais quantiques produisant un effet partiel, les effets à 0 et à 100 %. Pour de tels ensembles de données, la S.-K. est privilégiée à la méthode binomiale parce qu'elle permet de calculer des limites de confiance que l'on peut considérer comme valables.

Avec de « bons » ensembles de données, la méthode de S.-K. peut donner des réponses très semblables à celles de la régression probit, mais elle pourrait ne pas donner de réponses fiables dans certaines circonstances. Les comparaisons auxquelles on fait allusion dans les

§ 4.4.1 et 4.4.2 ont montré que cela se vérifiait parfois après équeutage comme sans équeutage. La méthode sans équeutage pourrait donner des réponses très curieuses avec des données modérément ou fortement erratiques (les deux exemples D des tableaux 2 et 3). La méthode avec équeutage donne parfois une meilleure estimation du paramètre de toxicité, mais dans les cas où les données sont peu nombreuses, elle n'a pas abouti à une estimation (exemples A à C du tableau 3).

La S.-K. sans équeutage donnera presque assurément une CE 50 anormale, peut-être sans limites de confiance, si les données ne comportent pas d'effets à 0 et à 100 %. On pourrait tenter une analyse employant un équeutage minimal si l'ensemble de données renfermait des effets tout à fait petits et grands ($\leq$ 20 %, $\geq$ 80 %) ainsi qu'un effet partiel central. Un équeutage de 20 % est susceptible de conduire à une estimation convenable de la CE 50 et de ses limites de confiance.

Apparemment, il ressort des exemples dont il a été question dans les § 4.4.1 et 4.4.2 que la façon la plus raisonnable d'utiliser la S.-K. est de faire une estimation sans équeutage et avec équeutage minimal, au taux choisi par le programme informatique. On devrait choisir entre les deux estimations (si elles sont différentes) en les comparant aux données brutes et à un graphique sur lequel on aura porté ces données brutes. Cela demande du jugement, mais cela semble inévitable. Les programmes informatiques ne comportent aucun test de la validité de l'estimation.

Les premières méthodes publiées par Environnement Canada mentionnaient la méthode de Spearman-Kärber, mais elles ne la recommandaient pour l'analyse (par ex. EC, 1992b). Cependant, dans CE (2001a), son emploi a été spécifié pour des ensembles de données quantiques ne renfermant qu'un effet partiel, inanalysable par la régression probit ou logit. Dans les méthodes plus récentes d'Environnement Canada (EC, 2004a, b, c), on autorise la S.-K. avec équeutage limité, en donnant des conseils de prudence à l'égard des ensembles de données ne comportant qu'un effet partiel. Les expérimentateurs devraient respecter les limites de la S.-K., notamment pour ce qui concerne les méthodes d'essai d'Environnement Canada. Une approche utile comprendrait l'utilisation judicieuse de l'équeutage de la façon conseillée dans l'alinéa précédent.

L'expérimentateur devrait vérifier soigneusement les opérations utilisées par tout programme employant la méthode de Spearman-Kärber. Les programmes offerts au moment d'écrire ces lignes permettent à l'expérimentateur de choisir entre aucun équeutage et l'équeutage. Nous recommandons d'utiliser ces deux options. Certains programmes ont autorisé l'utilisateur à préciser le taux d'équeutage (par ex. celui du MEEO [OMEE]). D'autres (TOXSTAT, CETIS) offrent une procédure « automatique » en vertu de laquelle le programme choisit le taux minimal d'équeutage satisfaisant. Nous recommandons cette option « automatique » ou « minimale ».

4.5.7 Méthode binomiale

Méthode mathématique connue, la méthode binomiale est actuellement offerte sous forme de progiciel commode pour l'analyse quantique dans un programme de Stephan *et al.* (1978) et elle est également modifiée pour la plate-forme Windows (OMEE, 1995). Nous la recommandons pour les nombreux ensembles de données dans lesquels une concentration entraîne un effet nul sur les organismes en expérience et où la concentration supérieure suivante provoque un effet à 100 %. On doit aussi l'utiliser pour un ensemble de données dans lesquelles se trouve un effet partiel, mais qui ne peuvent pas être analysées de façon satisfaisante par la méthode de Spearman-Kärber.

Les opérations mathématiques sont très simples. Quand aucun effet n'est partiel, la méthode binomiale pose, par approximation, que la CE 50 est la moyenne des logarithmes des deux concentrations causant les effets de 0 et de 100 %. Elle n'estime pas de limites de confiance, mais elle emploie les mêmes concentrations comme bornes d'un intervalle prudent (large) à l'intérieur duquel se trouve la CE 50. Les vraies limites de confiance se trouveraient probablement bien à l'intérieur de cette fourchette (voir ci-dessous).

Le calcul de base d'une CE 50 peut se faire facilement, sans programme informatique, par la moyenne des logarithmes des deux concentrations qui encadrent la CE 50. C'est la moyenne géométrique, que l'on peut également estimer en multipliant les valeurs arithmétiques des deux concentrations, puis en extrayant la racine carrée du produit, comme dans l'équation 3.

$$CE\ 50 = \sqrt{(C_{inf.})(C_{sup.})} \qquad (3)$$

Où :

$C_{inf.}$ = la valeur arithmétique de la concentration « inférieure » sans effet ;
C_{sup} = la valeur arithmétique de la concentration « supérieure » causant l'effet total.

L'intervalle à l'intérieur duquel on présume que se trouve la CE 50 est donné par les deux mêmes concentrations.

De fait, cette méthode binomiale est une simple *interpolation linéaire* sur une échelle logarithmique de la concentration. L'appellation *méthode binomiale* a été retenue pour respecter un usage ancien. L'appellation *interpolation linéaire* est réservée à une autre méthode (§ 4.5.9) pour éviter la confusion, parce qu'elle a été utilisée aux États-Unis pour une technique particulière, pas toujours satisfaisante.

La méthode binomiale est très utile, comme nous l'avons prouvé avec les données du tableau 3, parce qu'il est fréquent de ne pas disposer d'effets partiels quand on soumet des effluents industriels à des essais. Si les concentrations étaient convenablement rapprochées, on ne devrait pas considérer l'essai produisant de telles données comme déficient, mais, plutôt, comme une réponse valable, nette et uniforme des organismes en expérience. Cela peut être le signe d'un essai très précis, comme en a discuté Stephan (1977) [34] et, en pareil cas, l'utilisation de la méthode binomiale est recommandée.

Dans les essais sans effet partiel, les vraies limites de confiance sont habituellement situées bien à l'intérieur de l'intervalle des concentrations causant un effet de 0 et de 100 %. Si on avait une gradation plus fine des concentrations, la limite inférieure pourrait être élevée, correspondant à la concentration causant un effet de 30 %, tandis que la limite supérieure pourrait basse, correspondant à la concentration causant un effet de 70 % (Doe, 1994) [35]. Cela a été prouvé dans le tableau 2, selon lequel les limites de la méthode binomiale étaient beaucoup plus prudentes (intervalle plus large) que les vraies limites de confiance de la méthode des probits.

La méthode binomiale est également recommandée si les données présentent un effet partiel, mais ne peuvent pas être analysées par la méthode de Spearman-Kärber, faute d'effet de 0 ou de 100 % ou pour d'autres motifs. Si la méthode des probits ou celle de Spearman-Kärber est valable, il n'est pas nécessaire d'appliquer la méthode binomiale et on *ne devrait pas* l'appliquer. Néanmoins, la méthode binomiale fonctionnera et elle approxime la CE 50 obtenue par des calculs plus sophistiqués. Les comparaisons du tableau 2 ont montré que les CE 50 estimées par la méthode binomiale étaient quelque peu plus fortes que celles qui avaient été estimées par la méthode des probits ou celle des logits.

4.5.8 Méthode graphique de Litchfield-Wilcoxon

Dans les décennies antérieures aux années 1970, cette méthode « de raccourci » graphique de la régression probit était bien utilisée à cause de la rareté des ordinateurs et des calculatrices scientifiques. Voici comment la méthode fonctionne : on commence par tracer à la main une régression, puis on vérifie la qualité de l'ajustement et on estime les limites de confiance par calculs simplifiés et nomographie (Litchfield et Wilcoxon, 1949). L'annexe L donne une description plus détaillée des techniques employées.

34. Stephan (1977) traite de la plupart des techniques d'estimation des paramètres quantiques de toxicité pour situer son programme informatique dans son contexte. Il justifie la méthode binomiale et celle de la moyenne mobile et il explique pourquoi les écotoxicologues ne devraient pas trop s'inquiéter quand ils n'obtiennent pas deux effets partiels dans les résultats d'un essai quantique. Ces effets étaient importants en pharmacologie, ce qui a donné naissance à la régression probit, parce que les chercheurs devaient s'assurer eux-mêmes des pentes des droites des probits avant d'estimer la puissance relative de deux substances. Stephan fait remarquer que, dans le type de travail toxicologique dont nous discutons ici, l'on peut obtenir des paramètres utiles de toxicité sans aucun effet partiel.

35. Les limites de confiance d'un essai avec un effet partiel peuvent se lire à partir des tables fournies par van der Hoeven (1991), mais seulement dans des circonstances exceptionnelles. Le rapport des concentrations expérimentales successives doit égaler 2. Il ne doit pas y avoir d'effet à la concentration immédiatement inférieure à celle l'effet partiel et, à la concentration immédiatement supérieure à celle de l'effet partiel, l'effet doit être de 100 %. Dans d'autres situations, il faudrait employer une « méthode numérique assez compliquée ».

Nous ne préconisons pas la méthode pour des estimations définitives, mais elle peut encore être utile pour la vérification des estimations de la CE 50 et de ses limites de confiance produites par un programme informatique. En effet, comme première étape de toute analyse, nous recommandons de tracer à la main une première droite, pour vérifier si les estimations de l'ordinateur sont convenables (§ 4.2.2).

La méthode de Litchfield-Wilcoxon pourrait aussi avoir comme utilité d'aider à la formation du nouveau personnel. Les étapes de la méthode graphique pourraient donner un aperçu de la façon dont les paramètres de toxicité et leurs limites de confiance subissent l'influence de divers types de données. Cela pourrait aider à reconnaître les résultats anormaux d'un programme informatique.

La méthode a déjà été utile aux estimations initiales sur le terrain, quand l'accès aux programmes informatiques était impossible. Désormais, les ordinateurs portables remédient à cette situation.

4.5.9 Interpolation linéaire

Bien que l'expression « interpolation linéaire » désigne une technique ordinaire et largement utilisée, nous la mentionnons ici en tant que méthode séparée parce que l'USEPA l'a désignée comme catégorie distincte, dotée en propre d'une méthode statistique distincte (USEPA et USACE, 1994) [36]. On l'appelle parfois « méthode graphique » (USEPA, 2000a). Si on concède provisoirement que la traduction anglaise de ces expressions s'applique à la procédure états-unienne particulière, on peut affirmer que celle-ci ne procure aucun avantage particulier, et nous ne la recommandons pas. Les méthodes recommandées, pour les besoins d'Environnement Canada, seraient les méthodes des probits ou des logits, celle de Spearman-Kärber ou la binomiale, selon le nombre d'effets partiels. La méthode états-unienne d'interpolation linéaire est l'équivalent exact d'une estimation par la méthode binomiale si le nombre de concentrations n'est que de deux, l'une donnant un effet inférieur à 50 % et l'autre supérieur.

L'interpolation linéaire (et la méthode binomiale) reposent sur l'hypothèse d'un changement linéaire d'effet en fonction du logarithme de la concentration. Pour les données sans effet partiel, l'une ou l'autre méthode équivaut à tracer une droite sur un graphique entre les logarithmes des concentrations causant des effets de 0 et de 100 %, puis à interpoler le logarithme de la concentration causant un effet de 50 %. Cela revient à calculer la moyenne selon la formule de l'équation 3.

Une autre raison pour éviter la « méthode d'interpolation linéaire » est que certains programmes informatiques ont été construits pour employer les valeurs arithmétiques de la concentration, d'où des estimations erronées. L'erreur a été corrigée, et les logarithmes ont été utilisés par l'USEPA et l'USACE (1994).

La « méthode d'interpolation linéaire » est exposée plus en détail dans l'annexe L. On y trouve une méthode plus générale d'interpolation linéaire, qui accepterait les ensembles de données comportant des effets partiels. Cette méthode pourrait, en principe, être utile dans une situation inhabituelle.

4.5.10 Méthode de la moyenne mobile

Cette méthode n'est pas recommandée pour les programmes d'Environnement Canada, mais, ailleurs, on l'a considérée comme une solution possible pour l'analyse des données quantiques, et Stephan (1977) l'a considérée comme « la méthode de choix » en toxicologie des organismes aquatiques. La méthode a permis d'estimer la CE 50 et ses limites de confiance de façon identique ou semblable à la méthode des probits appliquée aux « bonnes » données du tableau 2 (§ 4.4.1). Cependant, avec des données irrégulières, elle a donné des estimations anormales par rapport à celles d'autres méthodes (exemples D des tableaux 2 et 3).

La méthode, mise au point par Thompson (1947), a besoin des résultats d'au moins quatre traitements, entre lesquels les intervalles géométriques ou logarithmiques doivent être égaux. En outre, elle pose que la distribution des données est symétrique. Elle peut estimer la CE 50, mais non la CE d'autres pourcentages d'effet, comme la CE 25.

En théorie, avec la méthode de la moyenne mobile on devrait estimer la CE 50 avec un ou sans effet partiel, bien que, dans les exemples du § 4.4, elle ne donne pas

36. Accessible à l'adresse électronique suivante : http://www.epa.gov/nerleerd/stat2.htm.

les limites de confiance sans au moins un effet partiel. Dans la pratique, le programme ordinaire offert pour appliquer la méthode (Stephan *et al.,* 1978 ; OMEE, 1995) ne fonctionne pas à moins qu'il n'y ait deux effets partiels ou plus. La régression probit ou logit fonctionnerait avec les mêmes données, et nous la recommandons. La méthode de la moyenne mobile souffre de certaines limitations, décrites par Finney (1978), qui fait observer que ses carences inhérentes sont à peine contrebalancées par sa simplicité de calcul dans une époque où le traitement informatique est si bon marché. Peut-être la régression probit ne conviendrait-elle pas à certains ensembles inhabituels de données et que la méthode de la moyenne mobile permettrait de les analyser.

Pour un ensemble donné d'observations, la méthode de la moyenne mobile estime plusieurs ensembles de CE 50 et de leurs limites de confiance, un ensemble pour chaque « fourchette » utilisée dans les calculs, c'est-à-dire le nombre d'intervalles entre les concentrations prises en charge dans les calculs. Le programme de Stephan imprime les résultats des calculs utilisant plusieurs de ces fourchettes, de sorte que l'expérimentateur peut examiner les modifications produites par diverses fourchettes. La plus appropriée est signalée par la plus petite valeur de *g*, qui est imprimée par le programme de Stephan. La version du MEEO (OMEE) du programme de Stephan choisit la fourchette la plus appropriée et désigne celle qui a été utilisée. Finney (1978) conseille la fourchette la plus étendue possible, sans définir ce *possible*.

4.6 Évaluation de nouveaux programmes informatiques

Les données des tableaux 2 et 3 pourraient servir à évaluer de nouveaux programmes informatiques d'estimation de CE p. On pourrait comparer les résultats à ceux qui figurent dans les tableaux, particulièrement le tableau 3, pour les données manquant d'effets partiels. Dans le doute au sujet de l'utilité d'un nouveau programme, on pourrait analyser d'autres ensembles de données moins que parfaits avec le nouveau programme et en comparer les résultats avec ceux d'un programme plus puissant tels que SAS ou SPSS.

Les critères pouvant servir à l'évaluation d'un programme informatique d'analyse des résultats d'un essai de toxicité ont été énumérés par Atkinson (1999, légèrement remaniés), après examen des programmes disponibles.

- Absence de formats non appropriés de codage et de rapport (par ex. les spécifications de l'USEPA).
- Exigences et coût en matière d'équipement et logiciels.
- Nécessité d'acheter des programmes supplémentaires (par ex. EXCEL).
- Qualité et « convivialité » des instructions.
- Contraintes sur les analyses et la saisie des données.
- Limites imposées au nombre de concentrations et de répétitions.
- Méthodes incluses pour le calcul du paramètre voulu de toxicité (par ex. test logistique, test de Williams).
- Méthodes contraires aux recommandations d'Environnement Canada.
- Pertinence des réglages par défaut.
- Utilisation d'une échelle logarithmique de la concentration appropriée aux calculs.
- Traitement des nombres inégaux de répétitions.
- Ajustement inapproprié des témoins pour les essais quantiques.
- Existence et utilité des présentations graphiques de données.
- Inclusion de tests pour la qualité de l'ajustement.
- Disponibilité de statistiques sommaires et de tests simples.
- Limites de confiance correctes (par rapport à celles d'autres méthodes).

Tous ces critères pourraient ne pas s'appliquer à un programme particulier, mais ils constituent une ossature partielle pour l'évaluation. Le dernier point pourrait être quelque peu élargi. Bien que l'on suppose que le paramètre estimé de toxicité sera juste, de même que ses limites de confiance, on devrait en vérifier la

justesse en les comparant aux résultats de programmes acceptés et en examinant la représentation graphique des données. Dans le § 4.2.3 (note 27), nous avons relaté qu'un laboratoire avait constaté que les résultats d'un programme informatique nouvellement acheté étaient erronés après qu'on les eut comparés à des graphiques tracés à la main (K.G. Doe, EC, Moncton, N.-B., communication personnelle).

4.7 Méthodes non linéaires et autres méthodes possibles de l'avenir

Pour l'avenir immédiat, la régression probit ou logit semble probablement la méthode de choix pour estimer les CE p dans les essais de toxicité dont le plan d'expérience est classique. Cependant, de nouvelles approches telles que les modèles non linéaires sont mises au point pour l'analyse des données quantiques. Dans une certaine mesure, les nouvelles méthodes quantiques se présentent comme des extensions de méthodes élaborées pour les régressions de données quantitatives. Un exemple approprié est l'adoption d'une méthode de régression linéaire ou non linéaire par Environnement Canada (§ 6.5.8).

Quelles que soient les méthodes mises au point, elles doivent permettre l'estimation de la CE 50 et de ses limites de confiance, si les essais sont effectués dans le cadre des programmes de surveillance d'Environnement Canada. Les bons programmes informatiques produiront également une description de la courbe ajustée (telle que la pente s'il s'agit d'une droite) et ils mesureront la qualité de l'ajustement.

Une méthode est déjà disponible. C'est un programme complet d'analyse des données sur la toxicité offert par Kooijman et Bedaux (1996). Il s'agit principalement d'une régression non linéaire mais qui, dit-on, permet d'analyser les données quantiques sur la mortalité (CL 50), les concentrations efficaces (CE 50) et les temps efficaces (TE 50), dans tous les cas avec leurs limites de confiance (voir le § 5.1).

Des statisticiens utilisent individuellement des modèles non linéaires depuis nombre d'années d'années pour l'analyse de données quantiques et la détermination de la CL 50. L'approche a été décrite par Kerr et Meador (1996), et nous en discutons dans l'annexe M.

Des modèles linéaires généralisés pourraient ne pas convenir aux essais de toxicité en routine, qui produisent souvent des données avec un seul ou aucun effet partiel. Les modèles peuvent utiliser des effets nul et total, mais ils sembleraient compter fortement sur les effets partiels. Les modèles non linéaires sont discutés plus en profondeur dans les § 6.5.2 à 6.5.13.

Dans la section 5, nous discutons de méthodes supplémentaires, dont l'intérêt est moins immédiat, notamment le temps correspondant à un effet de 50 %, et de l'emploi d'un modèle des taux de mortalité, cette dernière technique étant probablement plus intéressante pour la recherche. Dans l'annexe M, il est question d'autres méthodes potentielles.

Temps efficaces, courbes de toxicité et analyse de la survie

Tous les sujets abordés dans la présente section concernent le temps pris par la matière toxique pour agir sur les organismes. Actuellement, la priorité des programmes d'essais d'Environnement Canada ne va pas à l'estimation des temps ; cependant, de telles méthodes possèdent des avantages qui permettraient d'affiner l'analyse ou qui pourraient les faire adopter.

5.1 Temps efficaces 50

Repères

- *Une méthode de rechange consiste à estimer le temps nécessaire pour agir sur la moitié des organismes à chacune des concentrations fixes faisant partie d'une série de dilutions. Les temps efficaces 50 % (temps efficaces médians, temps efficaces 50, TE 50) et leur modélisation peuvent être plus instructifs, plus éclairants et plus utiles, dans certaines situations exceptionnelles telles que de courtes expositions.*
- *Les essais visant à estimer le TE 50 procurent également les données permettant d'estimer la CE 50, si on a bien choisi les concentrations.*
- *Il n'existe pas de progiciel simple et pratique pour estimer les TE 50, mais il serait utile d'en mettre un au point.*

Au cours des dernières décennies, on a peu utilisé les temps létaux 50 (TL 50), mais, par le passé, c'était la façon ordinaire d'étudier l'écotoxicité. Bliss (1937) a utilisé des séries chronologiques logarithmiques pour prouver que les transformations log-probit étaient utiles dans les essais de toxicité létale. Le TL 50 était le paramètre de toxicité estimé dans les études des effets des pesticides sur des insectes (Finney, 1971) et dans les travaux canadiens classiques sur la tolérance des poissons et des invertébrés aquatiques aux températures létales, au manque d'oxygène, à la salinité et aux toxiques (par ex. Fry, 1947 ; Shepard, 1955 ; McLeese, 1956). Une méthode axée sur l'estimation des temps (par ex. le TL 50) pourrait aider à évaluer les effets rapides d'un toxique dangereux. Par exemple, elle permettrait de prévoir les effets néfastes potentiels pour le poisson franchissant le panache d'un effluent.

Pour déterminer les TE 50 aigus, on utilise, dans un essai de toxicité, un groupe d'organismes que l'on expose à plusieurs concentrations constituant une suite logarithmique ordinaire. On observe à des moments successifs, formant une suite logarithmique approximative, le nombre d'organismes touchés à chaque concentration. Pour le poisson, les temps d'observation pourraient être de 0,5, 1, 2, 4, 8, 14 ± 2, 24, 48, 96 heures et peut-être de 7 jours. Pour les organismes de moindre longévité, on pourrait ajuster l'échelle de temps vers le bas, comme il convient.

Pour une concentration donnée, on porte le pourcentage d'effet cumulatif sur une échelle probit en fonction du logarithme du temps d'exposition. On ajuste une droite à vue d'œil et on lit le TL 50 sur le graphique. Les droites tracées à toutes les concentrations et réunies pourraient être semblables à celles de la fig. 11, qui montre un exemple classique des temps de mortalité du poisson exposé à une teneur réduite en oxygène (Shepard, 1955). Dans une fourchette convenable, les fortes concentrations entraîneraient des TE 50 courts, et certaines concentrations faibles ne pourraient se traduire que par des mortalités de moins de 50 % (partie droite de la fig. 11).

La technique pourrait servir à estimer la toxicité sublétale, mais l'effet devrait être facilement observé et être immédiatement évident, non retardé. L'effet devrait être quantique ou, sinon, être défini par rapport à un témoin, de la même manière qu'une concentration inhibitrice (CI p). L'expression *temps efficace 50* (TE 50) convient aux effets sublétaux comme aux effets létaux.

On pourrait se servir d'une série de TE 50 pour tracer des courbes de toxicité comme celles de la fig. 12. À première vue, les courbes ressemblent aux courbes de toxicité habituelles (§ 5.2), mais les coordonnées sont

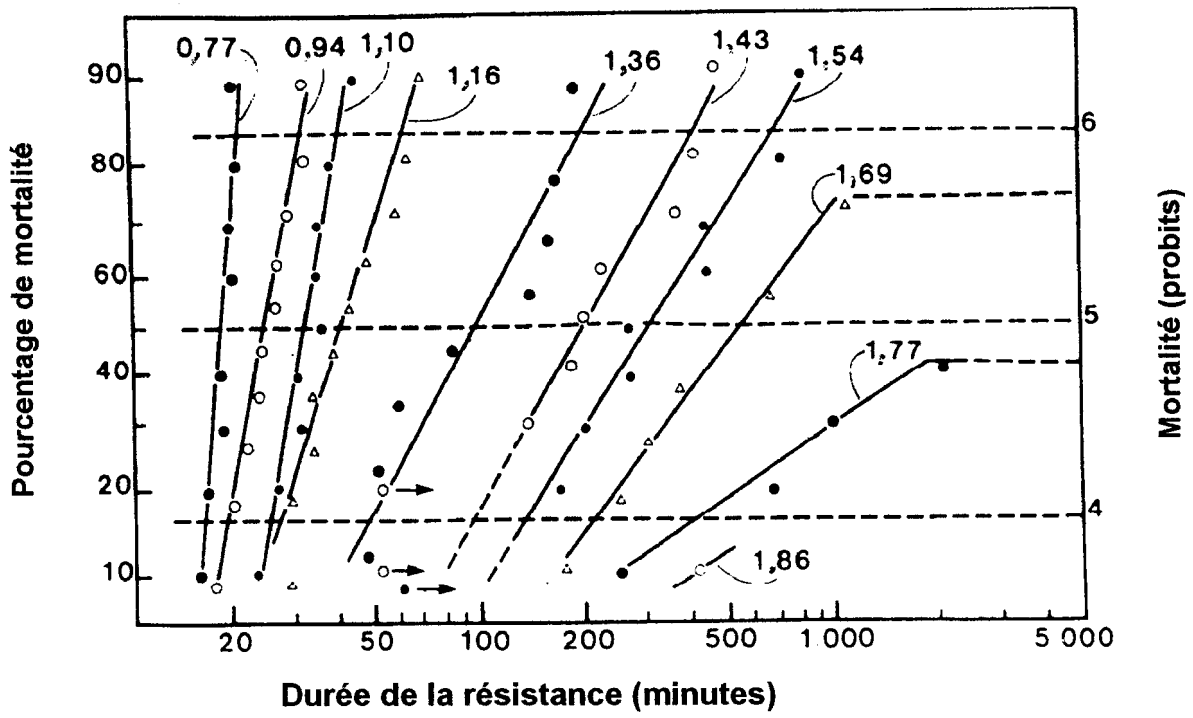


Figure 11. — Mortalité, en fonction du temps, de l'omble de fontaine exposée à de faibles concentrations d'oxygène dissous (d'après Shepard, 1955). Les concentrations d'oxygène figurent à l'extrémité supérieure des droites des probits. La durée maximale d'exposition de 5 000 minutes correspond à environ 83 heures. Les mortalités cumulatives successives de chaque groupe de poissons sont portées sur l'échelle verticale de probabilité et, à chacune, on ajuste une droite. La mortalité s'est apparemment interrompue dans les trois traitements les plus doux (à droite).

différentes : la concentration est en abscisse et le temps (TE 50) est en ordonnée. Les courbes de la fig. 12 tracées pour le cuivre et le zinc semblent plus droites qu'à l'accoutumée, avec des seuils d'effet très brusques. Sous ces concentrations seuils (côté gauche du graphique), plus de la moitié des organismes ont longtemps survécu ; apparemment, la toxicité létale aiguë a cessé d'agir, et les organismes ont pu s'adapter aux métaux.

Malheureusement, il n'existe pas de programme informatique simple, particulièrement conçu pour estimer les limites de confiance d'un TE 50 [37]. Anciennement, on obtenait ces limites de confiance par une méthode nomographique simplifiée (Litchfield, 1949). Les programmes informatiques ordinaires d'estimation de la CE 50 et de ses limites de confiance ne permettent pas d'estimer le TE 50, qui se déduit d'observations répétées sur les mêmes groupes d'organismes. Kooijman et Bedaux (1996) offrent un programme d'analyse des données sur la toxicité qui pourraient remédier à cette situation. Il sert principalement à l'analyse non linéaire des données quantitatives sur la toxicité sublétale, mais les auteurs allèguent qu'il permet également d'estimer des CE 50 et des TE 50 *avec leurs limites de confiance.* Ces capacités n'ont pas été vérifiées pour les besoins du présent document, en raison de difficultés initiales dans

37. Un programme informatique a été rédigé et utilisé à B.C. Research, quelque part dans les années 1970. Il dérivait de la méthode de Litchfield (1949) et il aurait bien fonctionné bien (D.J. McLeay, McLeay Environmental Ltd., Victoria, communication personnelle, 2004). Récemment, on a cherché, mais en vain, à retrouver ce programme.

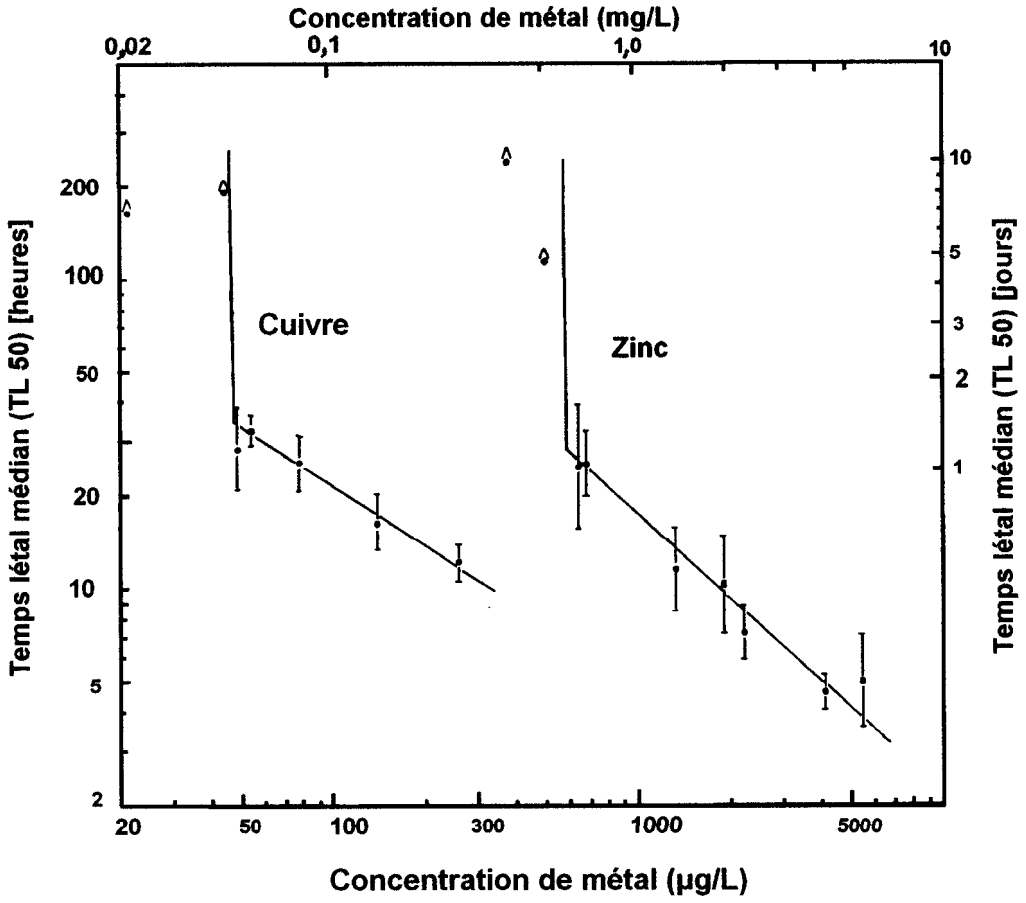


Figure 12. — Temps d'effet médian chez le saumon de l'Atlantique exposé au cuivre et au zinc (d'après Sprague, 1964). On a estimé les limites de confiance des TE 50 par la méthode de Litchfield (1949). Les points avec les flèches verticales représentent une survie de plus de 50 % des poissons en expérience durant la période d'exposition indiquée par la position sur l'axe du temps.

le fonctionnement du programme. Les progiciels usuels de statistique (SAS, SPSS, SYSTAT) pourraient estimer assez facilement le TE 50 et ses limites de confiance, bien qu'ils ne soient pas, à proprement parler, disponibles immédiatement pour un laboratoire de toxicologie.

L'emploi du TE 50 comme paramètre de la toxicité de chaque concentration est, d'une manière prévisible, plus efficacement instructif que la CE 50. En général, quand on estime seulement la CE 50, on perd la moitié de l'information. Dixon et Newman (1991) déclarent que des avantages statistiques considérables découlent du peu de travail supplémentaire à consacrer à l'obtention de données sur les temps de survie, par rapport à la détermination de la CL 50. De même, Newman et Aplin (1992) expriment le regret du peu de cas fait des méthodes d'estimation des temps efficaces en écotoxicologie. Ils font remarquer que cette approche n'empêche pas celle du paramètre ordinaire de toxicité (la CE 50), mais qu'elle permet d'obtenir des renseignements supplémentaires (la série de TE 50) et qu'elle aide à l'interprétation de données (grâce aux irrégularités significatives dans les effets).

Des exemples et d'autres explications sur le gain d'information sont donnés dans Bliss et Cattell, 1943 ; Gaddum, 1953 ; Sprague, 1969 ; Suter *et al.,* 1987. Un résultat attendu serait l'intervalle plus étroit de confiance pour le TE 50 par rapport à la CE 50. Un

autre avantage est d'éviter la complication entraînée par les estimations inversées de la CE 50 et de ses limites de confiance (v. le § 9.4).

Il y aurait encore davantage à gagner en renseignements, avec des méthodes qui tiendraient compte de la progression des effets (et non simplement du TE 50). Il pourrait y avoir des révélations supplémentaires sur ce qui se passe pendant un essai de toxicité. Parfois, on pourrait noter une pause dans la progression de l'effet, signe, peut-être, d'une modification dans le mécanisme de l'action toxique. Des différences entre les pentes des droites des probits adjacentes pourraient donner des indices sur les actions du toxique. Une interruption et l'aplatissement de la droite des probits pourraient signifier la décomposition de l'agent ou des agents toxiques actifs. Une double courbure de la ligne pourrait dénoter deux modes d'action à court et à long terme ou la présence de deux agents toxiques.

Un piège à éviter serait toute tentative de juger des toxicités relatives de différentes matières d'après les TE 50 à court terme (c'est-à-dire les temps efficaces de très fortes concentrations). La comparaison peut être très trompeuse (des exemples sont donnés dans Sprague, 1969). Induisent également en erreur les comparaisons de CE 50 fondées sur une courte exposition (de nouveau, mettant en cause de fortes concentrations). Les comparaisons sont beaucoup plus significatives quand elles se fondent sur des durées et des concentrations qui coïncident à peu près avec le seuil de l'effet (§ 5.2).

Compte tenu de tous les avantages des TE 50, il est regrettable que les méthodes aient tant privilégié l'estimation des CE 50 seulement. Une base de données rassemblées sur les TE 50 comme ceux des fig. 11 et 12 pourrait encore servir à l'estimation définitive de CE 50. On pourrait, par ex., estimer la CE 50 après 96 h de la manière habituelle, à partir du pourcentage d'effet aux diverses concentrations, après 96 heures d'exposition. On ne devrait utiliser que les observations brutes pour l'estimation de la CE 50 ; il ne serait pas valable de choisir les pourcentages d'effet lissés d'après les droites ajustées comme celles de la fig. 11.

5.2 Courbes de toxicité et seuils d'effet

L'expression *courbe de toxicité* a une signification particulière en écotoxicologie. C'est un graphique montrant une série de concentrations létales médianes (CL 50) tracées en fonction des durées d'exposition, les deux en logarithmes. Ce pourrait aussi être une série de temps létaux 50 (TL 50) en fonction des concentrations d'exposition, en logarithmes eux aussi (fig. 12).

Repères

- *On devrait tracer la courbe de toxicité à mesure que se déroule l'essai. On peut estimer les CL 50 à des moments cruciaux pendant l'essai et, à partir de leurs valeurs, tracer une courbe de toxicité (logarithme de la CL 50 en fonction du logarithme du temps).*
- *La courbe de toxicité révèle toute relation inhabituelle et elle montre si un seuil d'action toxique a été franchi avant la fin de l'essai (c'est-à-dire que la courbe devient asymptotique à l'axe du temps).*
- *La CL 50 initiale est un paramètre de toxicité relativement significatif, puisqu'elle est déterminée par la physiologie de l'organisme en expérience plutôt que par une valeur arbitraire de la durée d'exposition.*
- *La plupart des toxiques semblent produire une CL 50 initiale dans l'essai habituel d'exposition de poissons d'une durée de 96 h ainsi que dans les essais de toxicité d'un sédiment ou d'un sol pour des invertébrés durant de 10 à 14 jours.*
- *Si, en sus de signaler la CE 50 pour une durée standard d'exposition (par ex. la CE 50 pour le poisson, après 96 h), on signalait une CE 50 initiale ou l'absence de ce paramètre, on augmenterait la valeur pratique et scientifique de l'essai.*
- *La modélisation des données utilisées pour tracer des courbes de toxicité s'est révélée profitable dans les études de recherche (§ 5.3).*

La courbe a pour buts principaux de révéler toute relation inhabituelle et de montrer *si* on a atteint une

asymptote avec l'axe du temps. L'enregistrement périodique des effets pendant l'essai de toxicité aiguë permet de rassembler des données pour le tracé de la courbe de toxicité et augmente le gain d'information tiré de l'essai, ce qui est particulièrement vrai dans le cas des essais de toxicité létale aiguë. On prendra, comme exemples, ces essais employant des poissons[38].

Un objectif majeur de la courbe de toxicité est de déceler un seuil de mortalité indépendant du temps (c'est-à-dire cessation de la mortalité) et, le cas échéant, si ce seuil arrive tôt dans l'essai ou tard. On emploie le mot *seuil* dans le sens de moitié des poissons manifestant l'effet et l'autre moitié ne le manifestant pas, de sorte que le poisson médian a tout juste franchi le seuil d'effet (v. le glossaire). La concentration à laquelle survient ce phénomène peut s'appeler *CL 50 initiale* (ou *concentration létale initiale, CE 50 initiale, CL 50 seuil* ou *CE 50 seuil*). C'est une mesure relativement robuste de la toxicité puisqu'elle marque la concentration que le poisson moyen peut tout juste tolérer, en excrétant ou en détoxifiant une substance aussi rapidement qu'elle entre dans l'organisme. Autrement dit, la CL 50 initiale est déterminée par la physiologie du poisson ; c'est donc un paramètre descriptif de la toxicité aiguë qui est relativement significatif et sûr.

L'avantage de comparer les résultats correspondant à différentes durées d'exposition dans les essais de toxicité aiguë sont décrits par Sprague (1969), Newman et Aplin (1992) et Lloyd (1992). Si l'on ne trouve aucun seuil, c'est un avertissement que les effets pourraient continuer de se manifester pendant une exposition prolongée à de très faibles concentrations.

On peut estimer les CE 50 pendant le déroulement de l'essai (par ex. aux heures 4, 8, 24, 48 et 96 de l'exposition) et, grâce à elles, on peut tracer une courbe de toxicité sur des échelles logarithmiques (fig. 13)[39]. La courbe peut devenir manifestement asymptotique à l'axe de temps, c'est-à-dire que l'action létale aiguë a cessé (cas du toxique A dans la partie droite du graphique de la fig. 13). On a très fortement intérêt à savoir s'il existe une faible concentration que l'organisme moyen peut tolérer pendant une exposition aiguë ; les organismes survivants réchapperaient apparemment de l'exposition. Il n'existe pas de règle particulière permettant de déterminer si on a atteint une telle CL 50 initiale, de sorte qu'il faudrait interpréter subjectivement la courbe de toxicité[40]. Parfois, le seuil peut être très marqué, et son interprétation laisse peu de doute (fig. 12).

Même si une courte exposition n'a pas causé 50 % de mortalité, elle peut quand même contribuer à l'allure de la courbe de toxicité. Pour cette durée d'exposition, la CL 50 serait supérieure à la plus forte concentration expérimentale ; on peut placer un point accompagné d'une flèche pointant vers les concentrations supérieures à celles qui ont été expérimentées (extrémité gauche des courbes des fig. 13 et 14). La courbe ajustée peut ne pas englober certains points (lissage) parce que chaque CL 50 possède une variabilité (limites de confiance).

Il aurait été souhaitable de prolonger l'essai sur le toxique B de la fig. 13 pour voir si on pouvait finir par atteindre un seuil (asymptote). On devrait donc tracer grossièrement la courbe à mesure que l'essai avance, pour obtenir des indices sur la fin de l'essai. Même de faibles concentrations auraient apparemment tué les organismes, si l'exposition avait été prolongée. Il serait visiblement intéressant de connaître une telle situation, qui représenterait un type dangereux de toxique, parce que les concentrations de plus en plus faibles pourraient causer un effet, si la l'exposition était suffisamment longue.

L'emploi des logarithmes du temps et de la

38. Les essais de toxicité aiguë employant des poissons durent typiquement 4 j. Pour déterminer la mortalité aiguë d'invertébrés dans un sédiment ou un sol, les essais d'Environnement Canada durent d'habitude de 10 à 14 j, parfois avec inspection facultative de la mortalité à 7 j (EC, 1992e ; 1997a, b ; 1998b ; 2001a ; 2004a). Pour les essais sur un sol ou un sédiment, il est généralement impossible d'établir une courbe de toxicité en raison de la difficulté d'établir la mortalité à des moments intermédiaires et en raison du risque de blesser les animaux pendant l'inspection.

39. On utiliserait une échelle arithmétique de concentration au lieu d'une échelle logarithmique, si « l'agent toxique » à l'étude était la température ou le pH, lequel est déjà un logarithme.

40. Aucune méthode usuelle de test statistique n'a été établie pour déterminer si on a affaire à une asymptote. Il semble peu probable qu'une méthode simple sera accessible à cette fin, en partie en raison des observations non indépendantes répétées sur les mêmes groupes d'organismes.

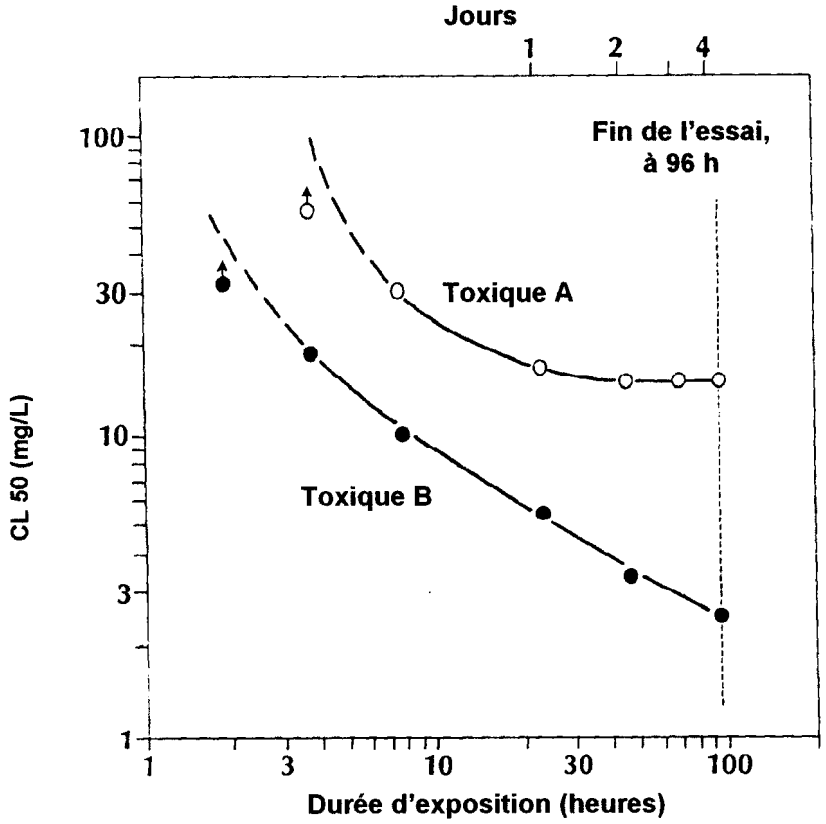


Figure 13. — Courbes de toxicité de deux toxiques hypothétiques. Ces courbes ont été ajustées à vue d'œil à toutes les CL 50. Des échelles logarithmiques sont utilisées pour le temps et la concentration. Le toxique A a atteint une CL 50 initiale, parce que la courbe devient asymptotique à l'axe du temps après deux jours environ. La courbe du toxique B n'est pas devenue asymptotique.

concentration, pour le traçage de la courbe de toxicité est d'une importance extrême, pour les raisons exposées dans le § 2.3 et l'annexe D. La courbe de toxicité tracée à une échelle arithmétique du temps est déformée et peut être fortement trompeuse. Une grave erreur pourrait être que le seuil a semblé être atteint à de longues expositions alors que, effectivement, il n'y aurait pas de seuil. *Avec un graphique utilisant une échelle arithmétique du temps, tout essai pourrait être amené à montrer un seuil apparent, même s'il n'en existe pas, simplement en le laissant courir assez longtemps.*

La fig. 14 montre un exemple hypothétique d'axes erronés : le graphique du haut emploie des échelles arithmétiques pour la concentration et le temps. La courbe semble atteindre une asymptote rassurante après 7 à 10 jours (168-240 h) d'exposition. Cependant, la représentation convenable des mêmes données, sur une échelle logarithmique, dans le graphique du bas, montre une mortalité régulière continue et une relation linéaire sans seuil. Autrement dit, si on utilisait des axes arithmétiques, on serait amené à croire, à tort, que le toxique possède un seuil, sous lequel l'effet toxique cesse quand, en fait, les faibles concentrations se sont révélées toxiques en vertu de la même relation, exactement, entre le temps et la concentration, qu'aux fortes concentrations.

Contribuerait à la mauvaise interprétation des données utilisées dans la fig. 14 le fait de ne pas augmenter l'exposition de façon régulière. Le changement important dans l'exposition est le *rapport* entre les concentrations successives d'exposition et non la valeur

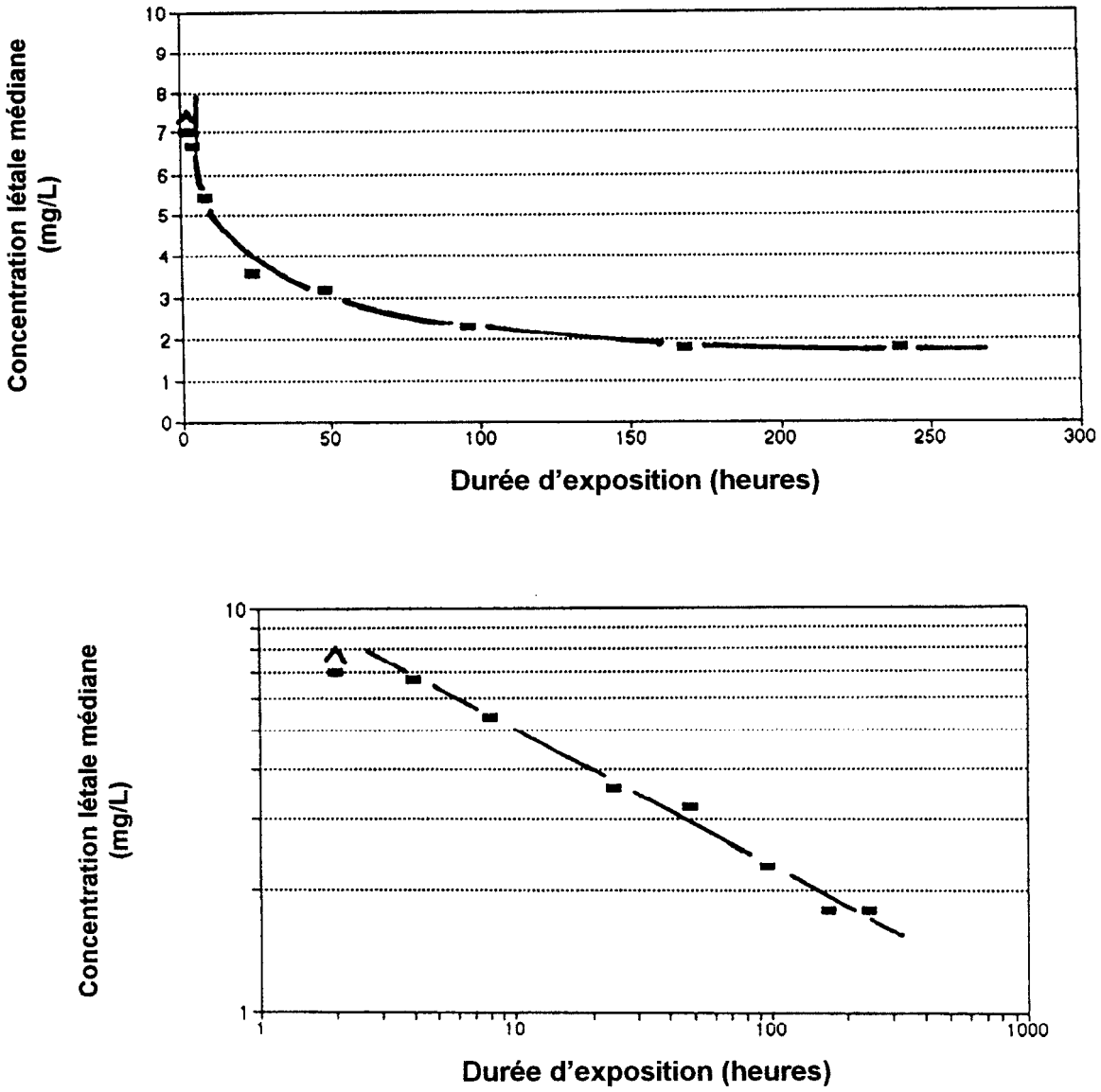


Figure 14. — Inadaptation de la courbe de toxicité sur un graphique employant des échelles arithmétiques. Dans le graphique du haut, les axes sont gradués selon une échelle arithmétique, et la courbe s'aplatit et devient parallèle à l'axe du temps vers la droite. L'expérimentateur serait induit à croire, mais à tort, qu'un seuil de toxicité aiguë a été franchi, de sorte que la toxicité ne se manifestera pas à des concentrations inférieures. Dans le graphique du bas, les données sont portées, cette fois, sur un graphique employant correctement des axes gradués selon l'échelle logarithmique, ce qui redresse la courbe de toxicité. On ne décèle aucun seuil, et la toxicité aiguë semble pouvoir continuer de se manifester à des concentrations plus basses, ce qui est une propriété d'un toxique dangereux. Les données sont hypothétiques.

absolue de l'augmentation (§ 2.3). Dans la plus grande partie de l'essai représenté dans la fig. 14, les durées successives d'exposition doublent ou presque. La paire finale d'inspections représente un intervalle de trois jours (du jour 7 au jour 10) qui pourrait sembler relativement long, ce qu'il est, en effet, à l'échelle arithmétique. Cependant, il représente une augmentation de seulement 1,4 fois le temps et, en

conséquence, autorise moins de changement de l'effet observé que les doublements antérieurs, par ex. des jours 1 à 2 et 2 à 4. Il faudrait remédier à ce type d'erreur, qui semble spontané dans les essais de toxicité aiguë d'un sol (Lanno *et al.,* 1997).

Estimation de la CE 50 initiale. — Sur une courbe de toxicité, il n'est pas approprié de signaler une CE 50 initiale ayant été estimée à vue d'œil. Au lieu de cela, la courbe sert à déterminer une *durée d'exposition qui semble se situer dans la région asymptotique,* et on calcule, pour cette durée, une CL 50 finale (initiale) à l'aide d'une technique usuelle (§ 4.5), qui donne une CE 50 exacte avec ses limites de confiance à 95 % .

L'OCDE (OECD, 2004) déconseille l'emploi des courbes de toxicité, estimant qu'elles ne constituent pas une méthode appropriée. Cependant l'argument ne convainc pas, et la principale objection statistique est que les données sur la relation dose-réponse à différents moments ne sont pas indépendantes. Cela ne semblerait pas faire problème, puisque la courbe de toxicité est simplement une façon de visualiser en toute simplicité le moment où des effets aigus semblent avoir cessé. Comme nous le recommandons, le calcul final de la CE 50 initiale se fait d'une manière usuelle, complètement indépendante de toute donnée sur les effets antérieurs. La courbe de toxicité peut jeter beaucoup d'éclairage sur les effets toxiques agissant dans un essai, et, à cause des observations de l'OCDE, il ne faudrait pas se priver de recourir à cet outil.

Établir le plan d'expérience d'un essai, en prévoyant l'établissement d'une courbe de toxicité, pourrait exiger la mise à l'essai d'un nombre supplémentaire de faibles concentrations. En contrepartie, cependant, la courbe de toxicité permettra habituellement de mieux comprendre le danger que pose le toxique. Dans les essais avec le poisson, la plupart des toxiques produisent une CL 50 initiale dans les 96 h usuelles (Sprague, 1969), tandis qu'un seuil semble probable dans les essais de toxicité d'un sol avec les vers de terre d'une durée de 14 jours (Lanno *et al.,* 1997).

Dans le cas des essais d'Environnement Canada, on devrait estimer une CE 50 pour la durée usuelle d'exposition stipulée dans le recueil de méthodes, par ex. 96 h pour les poissons ou 14 j pour les vers de terre. Si cette CE 50 « standard » représentait également une CE 50 initiale, comme nous l'avons décrite, il faudrait le signaler. Si on a obtenu une asymptote seulement après une exposition plus longue, on devrait estimer une deuxième CE 50 initiale pour cette période plus longue et la signaler comme paramètre supplémentaire et significatif de toxicité. Il est avantageux de présenter dans tout rapport sur un essai de toxicité létale une courbe de toxicité. Si on n'a observé aucune asymptote, on devrait le signaler ; *l'absence apparente de seuil est d'une importance toxicologique considérable.*

Il semble y avoir un regain d'intérêt dans la modélisation des effets toxiques en fonction du temps, comme on le montre dans les paragraphes qui suivent.

5.3 Modélisation des temps efficaces et courbes de toxicité

La modélisation statistique des résultats ne fait pas partie des essais normalisés d'Environnement Canada, de sorte que l'on ne fera qu'effleurer le sujet, au bénéfice des expérimentateurs qui pourraient souhaiter approfondir l'analyse de leurs résultats expérimentaux.

Quelques publications portent sur les descriptions statistiques des courbes de toxicité. Dans une étude innovatrice, Alderdice et Brett (1957) ont ajusté une hyperbole rectangulaire aux données sur la létalité d'un effluent d'usine canadienne de pâte à papier. Une CL 50 initiale a été calculée. Carter et Hubert (1984) ont produit une équation polynomiale généralisée (du type courbe de croissance), en utilisant un modèle linéaire multivarié. Hong *et al.* (1988) l'ont intégrée dans un programme informatique en langage BASIC. Ils ont utilisé le programme pour décrire un essai de toxicité de 14 jours avec des poissons et ils ont produit un graphique tridimensionnel des concentrations létales à p % (CL p) indépendantes du temps et des courbes de toxicité assorties de zones de confiance. Le programme n'a pas trouvé une grande utilisation. Il avait comme défaut de ne pas autoriser les effets observés chez les témoins et de modéliser les phénomènes avec les valeurs arithmétiques du temps, de sorte que les courbes donnaient une impression déformée des rapports de toxicité.

Heming *et al.* (1989) ont utilisé des analyses temporelles dans une excellente étude des effets de

l'insecticide méthoxychlore chez plusieurs espèces de poissons. Ils ont pu démontrer plusieurs ajustements pour les courbes de toxicité usuelles. Quatre modèles, sur les huit essayés, ont donné de bonnes descriptions des courbes. Kooijman et Bedaux (1996) offrent un programme complet (programme *DEBtox*) pour l'analyse des données sur la toxicité. Le programme possède des options pour l'analyse de données sur des CE 50 et des TE 50, avec leurs limites de confiance et la prise en considération du temps de réponse. D'autres ont utilisé le tracé d'un modèle de survie ajusté pour montrer une relation tridimensionnelle entre la concentration, le temps et le pourcentage d'effet (Newman et Aplin, 1992).

Périodiquement, on a tenté, il y a quelques décennies, d'extrapoler les courbes de toxicité létale afin de prévoir les effets toxiques de seuil, y compris les effets sublétaux. Lee *et al.* (1995) ont relancé cette quête, subtilement, en construisant trois modèles de prévision des effets létaux chroniques chez les poissons. Ils ont appliqué des régressions multiples aux données sur la toxicité létale aiguë, en transformant certaines données en logarithmes de la concentration et en logarithmes du temps, en réciproques du temps ou en logarithmes de la réciproque du temps. Des essais effectués avec 28 ensembles de données ont montré que les valeurs prévues étaient généralement proches des valeurs observées de la toxicité létale chronique et qu'elles étaient au moins du même ordre de grandeur. Dans la pratique, la méthode utiliserait des essais de toxicité aiguë bon marché pour identifier les polluants dangereux qui méritent d'être étudiés par des essais plus coûteux de toxicité chronique.

5.4 Analyses de la survie au fil du temps

Repères

- *Les taux de mortalité ou de survie et leur analyse représentent un groupe de méthodes statistiques avancées d'examen des effets toxiques. Ces méthodes sont bien connues en recherche biomédicale, et des publications récentes montrent leur pertinence pour l'écotoxicologie. Les techniques de recherche auraient besoin d'être adaptées à une utilisation en routine par les chercheurs.*
- *Les méthodes statistiques des mesures répétées pourraient souvent convenir à l'analyse des observations expérimentales répétitives.*

5.4.1 Taux de mortalité

La mortalité et la survie représentent les deux facettes de la même médaille, mais Borgmann (1994) a mis au point une méthode qui, en écotoxicologie, intègre les effets du temps et de la concentration sous l'appellation de *taux de mortalité.* La méthode pourrait être profitable à la recherche, particulièrement dans le cas de longues expositions combinant des observations sur la mortalité à des observations d'effets sublétaux, par ex. le poids. Elle serait utile à des essais à long terme sur les sédiments et employant des invertébrés, au cours desquels la mortalité est souvent un phénomène continu. Elle est également avantageuse lorsque l'on utilise peu de concentrations ayant des effets partiels. Les chercheurs intéressés pourraient mieux connaître les méthodes et leurs applications en consultant Borgmann (1994).

Bien que le taux de mortalité soit une variable continue ou quantitative, Borgmann (1994) l'utilise pour intégrer la mortalité, qui est un effet quantique. Le modèle du taux de mortalité part de l'hypothèse différente selon laquelle tous les organismes en expérience ont la même sensibilité à l'égard de la matière toxique et que la mortalité est un événement aléatoire que l'on peut quantifier comme un taux. Le taux de mortalité total peut être statistiquement séparé en taux de mortalité chez les témoins et en taux de mortalité causé par le toxique. On peut produire une courbe concentration-effet et estimer la CL 50. On peut également employer la méthode pour estimer la production de biomasse. Un manuel, de Fleiss (1981), donne des conseils sur la manipulation des taux.

5.4.2 Analyse de la survie

L'expression *analyse de la survie* englobe un groupe particulier de techniques, souvent utilisées dans les études biomédicales. Il s'agit de méthodes bien établies et profitables d'examen des effets toxiques par rapport au temps, bien qu'elles soient quelque peu complexes (Newman et Aplin, 1992). Crane et Godolphin (2000) en donnent une courte mais excellente introduction. Ils fournissent des exemples et citent des publications sur des sujets tels que la *régression linéaire en deux étapes,* la *régression probit multifactorielle,* la

modélisation du temps de survie et les *modèles cinétiques*. L'approche cinétique comprend la prise en considération plus ou moins théorique du comportement des toxiques dans les organismes vivants, avec la possibilité de mieux déterminer les concentrations toxiques initiales et les *vraies* concentrations sans effet (CSE).

Heming *et al.* (1989) ont appliqué ces techniques dans leur étude appliquée de la toxicité d'un pesticide (v. le § 5.3). Un autre bon exemple de modélisation du temps de survie est donné par Newman et Aplin (1992), qui ont analysé la toxicité du sel pour un poisson d'eau douce. Ils ont effectué des analyses usuelles des CL 50, mais ils ont montré que la modélisation du temps de survie était beaucoup plus instructive. Leurs méthodes ont permis prévoir les temps médians de survie à toute concentration donnée de toxique, les faibles taux de mortalité tels que 5 % et la toxicité pour une masse donnée du poisson, le tout accompagné d'estimations des erreurs types. Newman et Aplin (1992) ont recommandé le processus LIFEREG de la méthode SAS pour ces analyses.

Parmi les partisans les plus convaincus de ces analyses raffinées pour l'écotoxicologie se trouvent Kooijman et Bedaux (1996 ; également Kooijman, 1996). Une introduction exhaustive à ces sujets avancés, qui s'adresse à ceux qui possèdent un peu de compétences en statistique se trouve dans un livre récent de Crane *et al.* (2002). Le chapitre 5 montre les avantages de la modélisation du temps de survie par rapport aux analyses probit ou logit classiques de la toxicité létale aiguë. Le livre aborde des techniques plus avancées d'analyse temporelle, comme les tables de survie et les fonctions exponentielles de survie. Dixon et Newman (1991) font remarquer que les analyses des temps efficaces sont facilement mises en œuvre avec plusieurs progiciels courants, notamment SAS et SYSTAT, mais que ces progiciels ne représentant pas un programme facilement accessible et adapté aux besoins de tous les laboratoires de toxicologie. Une autre source d'information sur l'analyse de la survie est Parmar et Machin (1995).

5.4.3 Mesures répétées

On appelle *mesures répétées* les méthodes et les analyses fondées sur des mesures étalées dans le temps et provenant de la même source. Si un échantillon de sang était prélevé d'un poisson à plusieurs reprises, il donnerait des mesures répétées sur une *unité d'échantillonnage.* Si les mesures étaient effectuées, au fil du temps, sur des aliquotes d'une suspension d'algues extraite d'un plus gros récipient, les mesures répétées seraient faites sur l'*unité expérimentale.* (Ce ne serait pas des *sous-échantillons,* lesquels seraient prélevés simultanément.) L'approche n'est pas souvent utilisée en écotoxicologie, et les modifications de l'effet au fil du temps « peuvent et, souvent, devraient être analysées à l'aide de mesures répétées et de méthodes connexes, mais ces dernières risquent d'être plus complexes » que le plan d'expérience montré dans un tableau établi pour l'analyse de variance (Paine, 1996). On a besoin, en écotoxicologie, d'un modèle d'utilisation de ces approches plus sophistiquées aux données sur les effets en fonction du temps.

Estimations ponctuelles pour les essais quantitatifs de toxicité sublétale

L'estimation des paramètres de toxicité à la faveur d'essais de toxicité sublétale présente un intérêt majeur en écotoxicologie. Quatre des neuf sujets discutés par les écotoxicologues canadiens à la réunion de Québec portaient spécifiquement sur la détermination des paramètres de toxicité sublétale (Miller *et al.*, 1993).

La présente section débute par des conseils sur le choix des paramètres de toxicité et par des généralités concernant tous les essais de toxicité sublétale, puis elle passe aux points particuliers des *estimations ponctuelles quantitatives* que l'on peut utiliser pour décrire un effet sublétal. Les essais de toxicité sublétale sont traités de façon plus approfondie dans les sections 7 et 8.

6.1 Généralités sur les essais de toxicité sublétale

Dans un essai quantitatif de toxicité, l'expérimentateur n'observe pas simplement si l'organisme manifeste un effet ou non, mais, plutôt, il effectue des mesures quantitatives (en continu). Il pourrait mesurer le poids de chaque organisme en grammes, compter sa progéniture, mesurer l'activité d'une enzyme, etc. Les effets sur l'organisme entier sont d'un grand intérêt pratique. Nous les aborderons. Les effets généralement mesurés sont la taille atteinte, le degré de développement larvaire, la fécondation, la germination et le nombre de jeunes engendrés. Dans quelques cas, les effets sublétaux sont quantiques, mais on peut les assimiler à des effets quantitatifs en raison des nombreuses observations (v. le texte qui suit).

Les méthodes quantiques décrites dans les sections 4 et 5 ne sont ni appropriées ni valables pour les mesures quantitatives, et on ne devrait pas tenter de les appliquer. Cependant, la mortalité pourrait parfois être une mesure supplémentaire dans un essai conçu pour révéler les effets sublétaux, et l'analyse quantique serait appropriée pour ces données sur la mortalité dans des essais de mesure d'un double effet (section 8).

Les essais de toxicité sublétale disposent d'un choix d'approches et de méthodes, et nous formulerons des observations sur ce choix. Nous aborderons aussi certaines généralités, puisqu'elles s'appliquent aux deux estimations ponctuelles (la présente section) et aux tests d'hypothèses (section 7).

6.1.1 Types de tests et de paramètres de toxicité

Repères

- *Environnement Canada a publié diverses méthodes pour soumettre à des essais l'eau, les sédiments et les sols, en y exposant des organismes de façon chronique, subaiguë ou aiguë.*
- *La plupart des essais s'intéressent à des effets quantitatifs, par ex. la mesure du poids des organismes. On pourrait aussi mesurer des effets quantiques dans le même essai, comme la mortalité après une longue exposition ou la mortalité de la première génération de vers de terre.*
- *Le meilleur paramètre quantitatif de toxicité que l'on recommande est une estimation ponctuelle. C'est habituellement un degré spécifié de diminution des performances, par rapport au témoin, le plus souvent de 25 % dans les essais d'Environnement Canada. Un exemple serait la concentration associée à un poids inférieur de 25 % à celui du témoin.*
- *Pour effectuer des estimations ponctuelles, on a communément utilisé des méthodes insatisfaisantes d'analyse. La méthode d'interpolation est facile, mais elle néglige beaucoup de données. Des méthodes plus perfectionnées, qui font appel à la régression linéaire et non linéaire se répandent et sont désormais la norme dans les nouvelles méthodes d'Environnement Canada sur les essais de toxicité des sols. Les méthodes exigent que le personnel de*

laboratoire comprenne les jugements à poser dans le choix des modèles mathématiques appropriés.

- *Le test d'hypothèse est communément utilisé pour déterminer les concentrations exerçant des effets significatifs par rapport à ceux que présente le témoin. Cette méthode a de nombreux défauts, et elle ne sera désormais plus recommandée (v. la section 7).*

Types d'essais. — Ces dernières années, Environnement Canada a produit des méthodes normalisées pour un certain nombre d'essais de toxicité sublétale, la plupart employant des organismes aquatiques libres et des organismes vivant dans les sédiments. D'autres méthodes d'essai de toxicité d'un sédiment ou d'un sol sont en développement. Les essais sont énumérés dans l'annexe A, et nous les énumérons brièvement pour indiquer la large gamme d'organismes et d'effets sublétaux (EC, 1992a-f ; 1997a, b ; 1998a, b ; 1999b ; 2001a, b ; 2002a ; 2004a, b et 2007). Certains essais de toxicité emploient le test d'hypothèse pour l'analyse, mais, pour la plupart, on recommande des estimations ponctuelles quantitatives. Certains essais sont à double effet (section 8) et sont indiqués de la sorte dans la liste. Le second effet est souvent quantique, d'habitude la mort, qui, visiblement, n'est pas sublétale.

Organismes	Type d'essai
Bactérie luminescente marine *Vibrio fischeri*.	Inhibition sublétale des fonctions dans le milieu liquide, révélée par l'intensité de la luminescence.
	Inhibition sublétale des fonctions dans le sédiment
Algue verte dulçaquicole *Pseudokirchneriella subcapitata* [auparavant *Selenastrum capricornutum*]	Inhibition de la croissance et de la reproduction, révélée par le nombre de cellules
Plante dulçaquicole *Lemna minor*	Inhibition de la croissance
Plantes terrestres	Levée et croissance des plantes exposées aux contaminants du sol
Vers polychètes marins et estuariens	Inhibition de la croissance et mortalité dans le sédument (effet double)
Vers de terre, dans le sol	Comportement d'évitement. Effectifs et croissance de la progéniture. Mortalité dans la première génération (effet double).
Collemboles	Effectifs de la progéniture et mortalité dans la première génération (double effet)
Oursins, plats, etc.	Réussite de la fécondation après exposition initiale du sperme, poursuivie après addition d'œufs.
Crustacé dulçaquicole, la daphnie *Ceriodaphnia dubia*.	Nombre de jeunes engendrés et mortalité à long terme des adultes (effet double)
Crustacés (amphipodes) marins et esturiens	Comportement apparent d'évitement du sédiment, capacité de creuser des galeries et de s'enfouir de nouveau ; mortalité après 10 jours (effet double)
Amphipode dulçaquicole *Hyalella azteca*.	Croissance (gain de poids) et mortalité dans le sédiment après 14 jours d'exposition (effet double)
Larves dulçaquicoles de chironomes *Chironomus tentans* ou *C. riparius*	Croissance (gain de poids) et mortalité dans le sédiment après 14 jours d'exposition (effet double)
Cyprinidé dulçaquicole, le tête-de-boule	Croissance des larves venant d'éclore de ce poisson et leur mortalité (effet double)
Salmonidé dulçaquicole	Réussite du développement des embryons , des embryons et des alevins ou des embryons, des alevins et des jeunes poissons

Paramètres quantiques de toxicité. — Un essai conçu pour mesurer des effets sublétaux pourrait aussi avoir la mortalité comme effet parmi plusieurs autres. Il pourrait y avoir une mortalité à court terme à de fortes concentrations. L'exposition à long terme pourrait avoir divers effets sublétaux qui, au bout du compte, s'accumuleraient et causeraient la mort. L'analyse de la mortalité devrait être effectuée par régression probit ou par une autre méthode quantique (section 4).

Les essais quantiques de toxicité sublétale sont peu nombreux. L'un d'eux permet de mesurer l'évitement du sol contaminé par les vers de terre (EC, 2004a). L'analyse aboutirait à une CE p, par le même processus quantique que pour l'estimation de la CL 50 (section 4). Deux autres essais mesurent le succès de la fécondation avec des gamètes de truite arc-en-ciel (EC, 1998a) et d'oursins (EC, 1992f). L'effet est quantique, mais on peut appliquer une analyse de rechange aux oursins, comme il est décrit dans le texte qui suit.

Estimations quantitatives sur des données quantiques. — Si le nombre d'observations (organismes) quantiques est élevé, au moins 100 dans une répétition, il est acceptable d'analyser les données comme si elles étaient quantitatives. Un exemple serait l'essai de fécondation d'oursins (EC, 1992f), pour lequel on emploi de 100 à 200 œufs par récipient. Les œufs sont classés comme fécondés ou non fécondés, c'est-à-dire les données quantiques susmentionnées. En raison du grand nombre d'observations, cependant, la modification du pourcentage d'effet causée par un individu réagissant serait suffisamment petite pour que l'on puisse considérer ces données comme si elles représentaient une distribution continue [41]. Environnement Canada recommande d'estimer la CI p, un paramètre quantitatif de toxicité, dans l'essai employant des oursins. Le test d'hypothèse est une option supplémentaire, bien qu'il conserve tous les inconvénients énumérés dans le § 7.1.2. Un autre exemple se trouve dans les essais de croissance et (ou) de reproduction d'algues, dans lesquels la variable de base est le nombre de cellules, qui est quantique. Comme il peut y avoir des milliers ou des dizaines de milliers de cellules, la distribution des nombres peut être considérée comme continue, et l'essai est traité comme s'il était quantitatif.

D'autre part, l'essai d'Environnement Canada employant des salmonidés à leurs premiers stades n'emploie que 40 œufs par récipient, pour un total de 120 par traitement (EC, 1998a). Les résultats de l'essai sont quantiques (œufs viables ou non viables), et les paramètres de toxicité à estimer sont la concentration efficace 25 (CE 25) et la concentration efficace 50 (CE 50), qui conviennent à ces nombres d'individus. Les nombres se trouvant dans les récipients ne sont pas assez grands pour que l'on traite les données comme si elles étaient quantitatives.

Estimations ponctuelles quantitatives. — Le paramètre quantitatif préféré de toxicité dans les essais de toxicité sublétale est appelé *estimation ponctuelle,* ce qui est un point précis sur l'échelle continue de concentration (v. le § 6.2.2 pour connaître la liste des avantages que cela comporte). D'habitude on choisit le paramètre de toxicité pour représenter un certain degré de réduction des performances par rapport au témoin, par ex. 25 % de moins de progéniture que chez le témoin. La méthode pose donc fondamentalement comme hypothèse qu'il existe une relation dose-effet raisonnablement régulière pour servir à estimer le paramètre de toxicité.

L'emploi d'une estimation ponctuelle entraîne deux problèmes principaux :

- 1° La sélection d'un degré approprié de diminution des performances est clairement un choix subjectif du chercheur ou un choix résultant du consensus de la profession (est-ce que ça devrait être une diminution des performances de 25 ou de 10 %, comme cela est assez fréquent en Europe ?). Le choix du plus grand degré d'effet (25 %) fera imputer le résultat à la matière à l'étude et non pas simplement à une variation expérimentale. Un effet moindre (par ex. de 10 %) signifiera que le paramètre de toxicité est proche d'une concentration vraiment « inoffensive » (v. le glossaire et le § 6.2.4).

- 2° Les distributions de l'effet prennent diverses allures et, en conséquence, leur description exige divers modèles mathématiques. Cependant, de réels

41. On peut prévoir que les données quantiques (binaires) suivront une loi binomiale, et les analyses statistiques appropriées emploieront des méthodes pour cette distribution, comme le test du khi-deux. Cependant, la distribution de nombreuses observations vient à ressembler à une distribution normale. On introduit peu d'erreur ou de biais en utilisant des techniques statistiques quantitatives pour estimer un paramètre de toxicité.

Par exemple, si, dans une répétition de 10 œufs, 8 se révèlent fécondés, chaque œuf a influé sur 10 % des résultats (de 70 % d'effet total si cet œuf n'avait pas été fécondé, à 80 % d'effet total, s'il avait été fécondé. Ce saut de 10 % est abrupt et constitue un changement appréciable, révélateur de la nature quantique des données. Cependant, dans une répétition de 100 œufs, chaque œuf pourrait n'influer que de 1 % sur le résultat global, disons de 70 à 71 %. En pratique, cela représente un effet quantitatif.

progrès ont été effectués dans l'élaboration d'une approche normalisée, qui débute par la sélection d'un modèle approprié parmi un petit éventail de choix (§ 6.5.8).

L'expérimentateur qui envisage des méthodes appropriées d'analyse pourrait s'inspirer de l'organigramme de la fig. 15 en descendant par la gauche jusqu'à la case « Estimation ponctuelle » et y trouver deux choix généraux de méthode, décrits dans les § 6.4 et 6.5.

- Le premier choix est la méthode non paramétrique de ***lissage et d'interpolation.*** Cette méthode d'analyse, jadis usuelle, souffre cependant de plusieurs défauts et mérite d'être remplacée (§ 6.4.1).

- Le second choix est la ***régression, linéaire ou non linéaire,*** convenant à diverses distributions dose-effet. Des programmes statistiques polyvalents tels que SYSTAT peuvent servir dans une approche analytique standard, désormais adoptée par Environnement Canada (§ 6.5.8). Le programme toxicologique CETIS offre aussi des modèles mathématiques de régression non linéaire. Il faut toujours posséder certaines connaissances mathématiques pour choisir le modèle non linéaire approprié et appliquer le traitement mathématique.

La participation d'un statisticien à l'établissement du plan d'expérience et à l'analyse (§ 2.1) est particulièrement importante pour les estimations ponctuelles de paramètres sublétaux. Certaines des méthodes les plus raffinées, qui se trouvent à la fin de la section 6 exigent absolument l'obtention de conseils en statistique d'une personne compétente.

Test d'hypothèse. — Cette solution de rechange aux estimations ponctuelles a été couramment utilisée et elle est autorisée, mais elle n'est plus recommandée dans diverses nouvelles méthodes d'essai d'Environnement Canada. L'approche consiste à déterminer la concentration minimale ayant causé un effet statistiquement significatif dans l'essai (la CEMO ; droite de la fig. 15). On la décrit avec ses carences dans la section 7.

6.2 Rudiments des estimations ponctuelles de paramètres de toxicité sublétale

Repères

- *La concentration inhibitrice p (CI p) pour un pourcentage spécifié de réduction des performances est le paramètre usuel de toxicité des essais quantitatifs de toxicité sublétale. La valeur de* p *dans l'expression* CI p *est généralement de 25 ou de 20 % ou, parfois, de 10 %, en Europe. N'ayant aucune racine statistique elle est choisie d'après le jugement du biologiste.*

- *Les Européens et certains groupes des États-Unis appellent souvent ce paramètre la* concentration efficace à p % *(CE p), erreur trompeuse, puisque cela fait allusion aux essais de toxicité quantique dans lesquels une proportion spécifiée d'organismes présente un effet particulier.*

- *La CI p est avantageuses à de nombreux points de vue. C'est une concentration unique, ses limites de confiance sont calculables, et la variabilité des données ne devrait pas influer systématiquement sur sa valeur. Les inconvénients sont moins nombreux et mineurs.*

- *La répétition pourrait ne pas être exigée par la méthode d'essai, mais même une répétition modeste est avantageuse. Elle peut aider à distinguer entre :* a*) la variabilité à l'intérieur de l'essai ;* b*) l'écart par rapport au modèle choisi de la relation dose-effet. Des répétitions poussées sont nécessaires si on veut déterminer les paramètres de toxicité avec une régression non linéaire.*

- *Avant de passer à l'analyse mathématique formelle, on devrait tracer manuellement un graphique des résultats pour permettre l'examen visuel de la courbe dose-effet et permettre la détermination grossière d'un paramètre de toxicité afin de vérifier la justesse de l'estimation faite par l'ordinateur.*

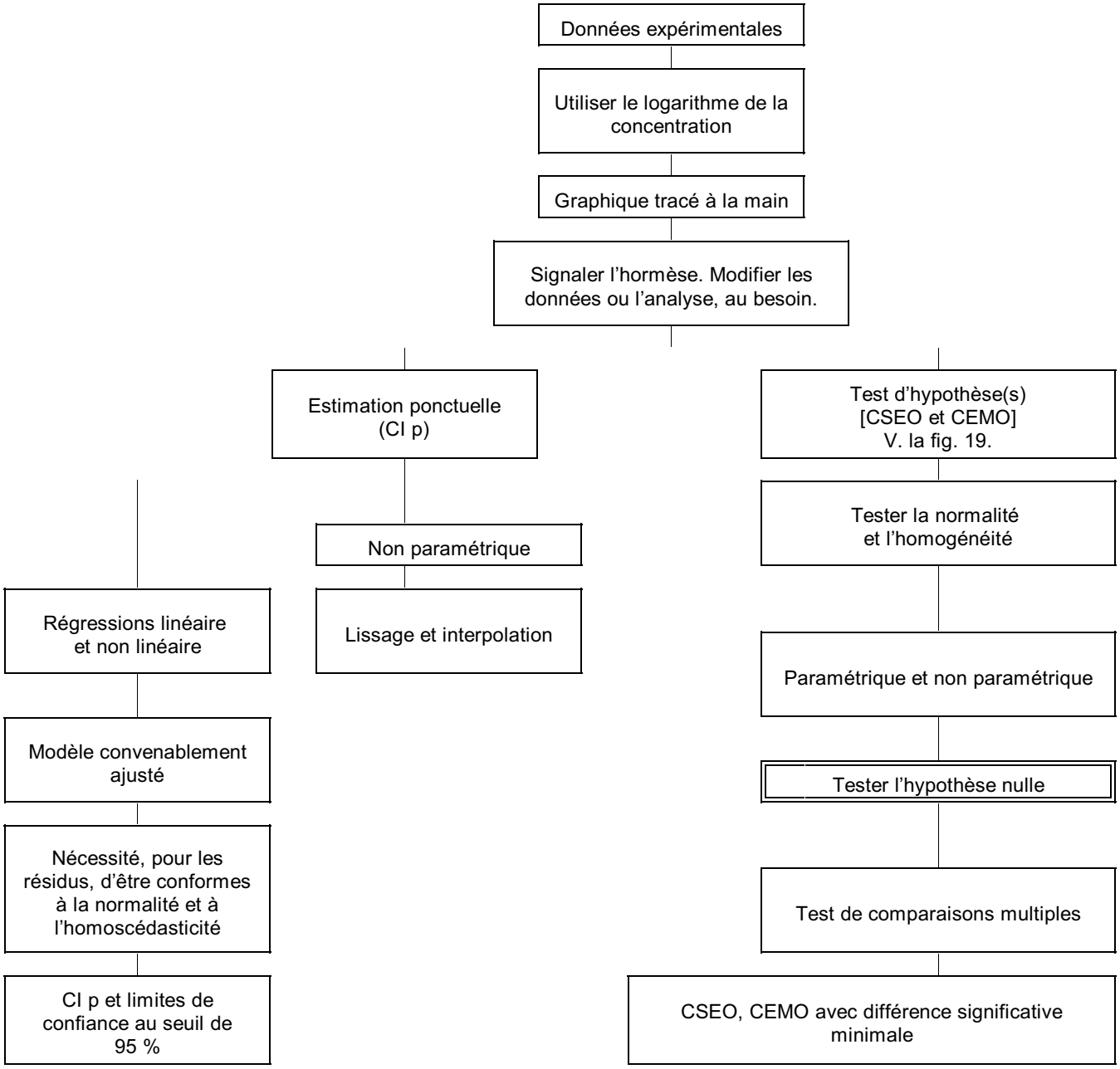

Données expérimentales
Utiliser le logarithme de la concentration
Graphique tracé à la main
Signaler l'hormèse. Modifier les données ou l'analyse, au besoin.
Estimation ponctuelle (CI p)
Test d'hypothèse(s) [CSEO et CEMO] V. la fig. 19.
Non paramétrique
Tester la normalité et l'homogénéité
Régressions linéaire et non linéaire
Lissage et interpolation
Paramétrique et non paramétrique
Modèle convenablement ajusté
Tester l'hypothèse nulle
Nécessité, pour les résidus, d'être conformes à la normalité et à l'homoscédasticité
Test de comparaisons multiples
CI p et limites de confiance au seuil de 95 %
CSEO, CEMO avec différence significative minimale

Figure 15. — Organigramme de l'analyse des résultats des essais de toxicité quantitatifs à plusieurs concentrations.

6.2.1 Terminologie

En Amérique du Nord, une estimation ponctuelle de la toxicité sublétale quantitative est la *CI p,* c'est-à-dire la *concentration inhibitrice à p %,* p % étant le pourcentage spécifié d'effet. C'est la concentration que l'on estime causer tel pourcentage de dysfonction biologique, par rapport au témoin. Par exemple, la CI 25 pourrait être la concentration que l'on estime réduire la progéniture de 25 % par rapport au témoin.

On ne devrait pas décrire les effets sublétaux quantitatifs par la CE 25, la CE 50, etc. ; ces expressions sont valables pour les données quantiques (la concentration efficace pour un pourcentage spécifié d'individus). 25 % d'individus touchés (CE 25), c'est tout à fait différent de performances diminuées de 25 % par rapport à celles du témoin (CI 25). La bonne terminologie informe sur le type d'essai, le type de données obtenues et le type approprié d'analyse. La mauvaise terminologie induit en erreur.

L'utilisation erronée de la notion de CE 50 en Europe, même par des techniciens réputés, de groupes de travail de l'OCDE et de l'ISO, est particulièrement inquiétante. Cette erreur se commet aussi en Amérique du Nord, dans certains progiciels de statistique (CETIS), chez des mathématiciens (ce qui est étonnant) et, notamment, dans les essais de luminescence bactérienne. Même l'USEPA, parfois, omet de distinguer nettement entre essais quantiques et essais quantitatifs, dans la description des estimations ponctuelles (USEPA, 1995).

D'autres termes et symboles ont été proposés pour les paramètres de toxicité estimés au moyen de techniques statistiques particulières, mais c'est la concentration inhibitrice p (CI p) qui semble convenir à toutes les estimations quantitatives.

6.2.2 Avantages des estimations ponctuelles

Les principaux avantages des estimations ponctuelles sont qu'une seule concentration simple est obtenue comme paramètre de toxicité et que l'on peut en estimer les limites de confiance. Nous énumérons ci-dessous d'autres avantages de cette méthode, par rapport à celle de la CSEO et de la CEMO. La plupart des avantages énumérés reposent sur l'hypothèse de l'obtention de l'estimation ponctuelle par régression. Des listes semblables ont été dressées par Stephan et Rogers (1985) ; Pack (1993) ; Noppert *et al.* (1994) [42] ; Chapman (1996) ; Moore (1996) ; OECD (1998) et d'autres.

a) Une seule concentration est désignée comme paramètre de toxicité.

b) Ce paramètre peut être n'importe quelle concentration située dans l'intervalle visé par l'essai et n'a pas besoin d'être une concentration choisie par l'expérimentateur et utilisée dans l'essai.

c) On peut accompagner le paramètre de limites de confiance. On peut calculer d'autres expressions usuelles de la variation telles que l'écart type.

d) La valeur du paramètre ne subirait habituellement pas l'effet d'une erreur systématique dans la même direction qui serait causée par le degré de variation naturelle, par la variation provoquée par la minutie de l'expérimentateur ou par le nombre de répétitions (la précision, toutefois, pourrait subir leur influence).

e) Le choix d'***α,*** le niveau de signification, ne modifie pas l'estimation de la toxicité.

f) Si le paramètre de toxicité est estimé par régression, la méthode choisie aurait d'habitude sur lui un effet relativement petit, au moins pour les valeurs centrales de *p* (de la CI p).

g) L'emploi de la CI p favorise la prise en considération des degrés d'altération dans le monde réel et dissuade de penser que la CSEO obtenue par un test d'hypothèse est une concentration « sans effet » biologique.

On peut aussi énumérer les inconvénients des estimations ponctuelles. Certains sont simplement des problèmes à résoudre ou des méthodes à normaliser.

a) L'ampleur de l'effet correspondant au paramètre de toxicité (la valeur de *p* dans *CI p*) n'est pas un absolu et elle exige un apport subjectif et l'accord entre les chercheurs.

b) La précision de l'estimation du paramètre dépend du nombre de concentrations expérimentées, de leurs valeurs numériques, du nombre de répétitions

42. Sinon, v. deBruijn et Hof (1997), van der Hoeven (1997) et van der Hoeven *et al.* (1997).

et du choix d'un modèle mathématique approprié pour décrire la relation. Ainsi, le choix des concentrations peut influencer sur la CI p estimée, particulièrement aux faibles valeurs de *p*.

c) Plus la valeur de *p* dans CI p diminue, plus l'intervalle de confiance s'élargit.

d) Le modèle choisi pour s'ajuster aux données peut influer sur la valeur estimée du paramètre de toxicité, particulièrement, de nouveau, si ce dernier correspond à un petit effet.

6.2.3 Répétitions

Le § 2.5 donne tous les renseignements relatifs aux répétitions pour les estimations ponctuelles. Pour la régression, une seule mesure à chaque concentration est la condition absolue permettant d'estimer le paramètre de toxicité et ses limites de confiance. Cependant, il faut des répétitions si l'on souhaite choisir parmi des modèles linéaires et non linéaires pour les ajuster aux données et évaluer la qualité de l'ajustement. Pour une estimation moins souhaitable du paramètre de toxicité par lissage et interpolation (programme ICPIN, § 6.4), il faut au moins deux répétitions à chaque concentration pour calculer les limites de confiance, et cinq ou plus sont souhaitables. Les documents d'Environnement Canada recommandent d'habitude trois répétitions pour les estimations ponctuelles, au cas où elles seraient nécessaires au test d'hypothèse, mais quatre seraient nécessaires à certaines méthodes d'analyse non paramétrique.

6.2.4 Choix du degré d'effet pour le paramètre de toxicité

Le choix d'une valeur de *p* (de la CI p) est entièrement arbitraire. C'est une décision de l'expérimentateur qui fait appel à son jugement. Les mathématiques n'ont rien à y voir. En Amérique du Nord, on a tenté, sans qu'il y ait rien d'officiel, d'établir la CI 20 comme paramètre standard de toxicité dans les essais en milieu aquatique, mais la CI 25 (c'est-à-dire réduction de 25 % des performances) est le plus souvent utilisée.

On aurait été justifié de croire que la CI 25 était semblable, dans des nombreux cas, à la CSEO [43]. L'argument n'est pas particulièrement convaincant, vu les nombreuses déficiences de l'approche fondée sur la CSEO et la CEMO (section 7). La relation entre la CI p et la CSEO pourrait changer selon la puissance et la variation d'un essai donné, l'effet mesuré et la matière soumise à l'essai.

Les Européens ont constaté qu'il était possible d'estimer la CI 10 dans plusieurs types d'essais. Cette concentration a servi à décrire l'inhibition de la croissance des algues (ISO, 1999) et elle est sanctionnée pour d'autres méthodes de l'organisation (ISO, 1998). La CI 10 est certes un paramètre acceptable de toxicité si on peut l'estimer avec un intervalle de confiance convenablement étroit et si on satisfait à certaines autres conditions (v. le texte qui suit). La promotion de la CI 10 comme éventuel paramètre de toxicité pourrait aider à remplacer les moins souhaitables CSEO et CEMO. Certains clients des programmes d'essais de l'industrie et d'organisations écologistes sont d'avis que le paramètre de toxicité estimé par un essai devrait sembler « inoffensif », ce qui le cas de la CSEO. La CI 10 est visiblement plus proche d'une concentration sans effet que ne l'est la CI 25 et elle donne l'impression plus nette d'être un paramètre de faible toxicité, assez rassurant.

Le Groupe consultatif sur la statistique d'Environnement Canada a énuméré divers facteurs influant sur le choix d'une valeur appropriée de *p* (Miller *et al.*, 1993). Les voici, sous forme lapidaire.

- La question fondamentale est de choisir la valeur de *p* (de la CI p) soit d'après son importance écologique, soit pour plaire aux statisticiens.

43. Les preuves ne sont pas nombreuses. Une comparaison importante ayant porté sur des essais en milieu aquatique a montré que la CI 25 était semblable à la CSEO dans le cas de 23 effluents et toxiques de référence, par suite d'essais de toxicité sublétale menés avec des oursins, le cyprinodon varié (*Cyprinodon variegatus,* un poisson) et l'algue rouge non microscopique *Champia* (USEPA, 1991a). Un ensemble d'essais avec des daphnies et un seul toxique de référence a également montré la similitude des deux mesures (OECD, 1997). Une compilation par Suter *et al.* (1987) de 176 essais de toxicité sublétale avec des poissons a montré que, en moyenne, la CI 25 était presque égale à la concentration avec effet de seuil observé (CESO), concentration plus forte que la CSEO. Cependant, même les rapports moyens de la CI 25 à la CSEO, pour ce qui concerne divers effets chez les poissons, ont varié de 0,5 à 3,2. Plus tard, Suter *et al.* (1995) ont conclu qu'une inhibition sublétale de 20 à 25 % était à peu près le minimum qui correspondrait à un effet statistiquement décelable (CEMO).

- Une faible valeur de *p* est souhaitable du point de vue biologique, pour obtenir une estimation sensible de l'action toxique.
- Une faible valeur telle que la CI 10 signifierait travailler à l'extrémité de la relation dose-effet, qui entraîne peut-être une variation indésirable de l'estimation. Une CI 50 serait statistiquement souhaitable, mais, biologiquement, une valeur inférieure de *p* serait exigée dans les essais de toxicité sublétale.
- L'option d'une faible valeur de *p* sera propre au type d'essai et à l'effet mesuré. Si l'effet mesuré est variable, la CI 10 pourrait bien se situer dans la zone de variabilité biologique normale, d'où une interprétation incertaine du paramètre de toxicité. La variabilité observée chez le témoin devrait influer sur le choix de la valeur de *p*.
- La CI 25 ou, parfois, la CI 20, semble avoir gagné la faveur en Amérique du Nord et d'autres pays, en tant que bon indicateur minimal d'une modification « biologiquement significative ».

La valeur de *p* (de la CI p) devrait être supérieure à toute valeur spécifiée dans les méthodes comme limite supérieure de l'effet acceptable chez le témoin, et cette précision devrait être ajoutée à l'avant-dernier alinéa de l'énumération qui précède. La CI 10 semblerait généralement une limite inférieure pratique comme paramètre fiable de toxicité. Les éléments à considérer dans le choix d'une valeur de *p* pour la CI p sont quelque peu analogues à ceux du choix d'un paramètre quantique de toxicité (la CE p), qui sont mentionnés dans le § 4.2.5.

Statistiquement, il est indéniable que l'intervalle de confiance de la CI p s'élargit à mesure que *p* diminue. Pour les valeurs très faibles de *p*, il pourrait être difficile d'obtenir une estimation convenable d'une concentration avec un intervalle de confiance suffisamment étroit. Une partie de cet effet est imputable à la possibilité que la CI p et ses limites soient estimées par *régression inverse,* à l'instar des estimations quantiques de la CE p (v. le § 9.4)[44]. Dans les régions extrêmes de la régression, l'intervalle de confiance s'élargit et, aux concentrations minimales, il n'est pas rare que la limite inférieure aille à l'infini quand on a employé la régression inverse.

Le choix des concentrations peut atténuer ce problème. Comme les intervalles de confiance commencent toujours à s'élargir à partir de la moyenne de la variable indépendante, un bon plan d'expérience ferait en sorte de centrer la variable indépendante (la concentration) sur la valeur de *p* à laquelle on s'intéresse. On devrait donc choisir les concentrations pour les rapprocher du paramètre envisagé de toxicité, disons la CI 10. Bien sûr, cette valeur est un peu difficile à prévoir, et les priorités s'opposent (§ 2.2), mais on devrait se rappeler ce principe.

L'expérience générale montre que la CI 10 est moins souhaitable dans les essais suivant généralement les méthodes d'Environnement Canada parce que ce paramètre peut posséder un large intervalle de confiance. La CI 25 est devenue usuelle à Environnement Canada et, plus généralement, en Amérique du Nord. Par ailleurs, la CI 20, sa solution de remplacement, a le mérite d'être un paramètre de toxicité que l'on peut estimer et qui reste significatif.

6.2.5 Sélection de la variable biologique comme paramètre de toxicité

L'*effet* à analyser pourrait influer sur la CI p, en la déplaçant vers le haut ou le bas. En conséquence, le choix d'effet pourrait influer beaucoup sur la sélection de la valeur de *p* de la CI p. Cela pourrait être particulièrement important dans des essais de mesure d'un double effet, sujet abordé dans la section 8.

44. En bref, on cherche un ensemble de concentrations et on observe les effets correspondant à chacune d'elles. Si une régression est ajustée aux données, l'effet est la variable dépendante, et le logarithme de la concentration est la variable indépendante. Les intervalles de confiance de la régression sont en fonction de l'*effet* et, comme toujours, ils sont plus larges aux concentrations extrêmes qu'au « centre ». On souhaite inverser l'interprétation, en exprimant les limites de confiance en fonction de la concentration autour d'un paramètre de toxicité (concentration) que l'on estime causer un effet indiqué (*p*). On peut considérer cela comme une « régression inverse ». Les limites inversées sont asymétriques, et, dans la région inférieure de la régression, l'intervalle inférieur de confiance peut s'élargir particulièrement (comme dans la fig. 7). V. le § 9.4.

6.3 Étapes générales de l'estimation d'un paramètre de toxicité sublétale

Repères

- *La première étape de l'analyse devrait être le tracé, à main levée, d'un graphique des données. Ce graphique montre la nature générale des résultats et permet de vérifier l'estimation finale du paramètre de toxicité.*
- *La méthode de choix est la régression linéaire ou non linéaire. Si elle est impraticable, on devrait se rabattre sur une méthode commune d'interpolation, le processus ICPIN.*

6.3.1 Tracé des données

Un tracé à la main devrait constituer la première étape de l'analyse et il n'a pas besoin de prendre beaucoup de temps. Tracer le graphique de l'effet en fonction du logarithme de la concentration, que l'effet soit la taille atteinte, le pourcentage de diminution de la reproduction ou un quelque autre effet quantitatif (v. un échantillon des résultats dans les fig. 22 à 31 de la section 10 et dans la fig. P.1 de l'annexe P).

Voici une liste des avantages que présente le graphique tracé à la main.

a) Le graphique révélera tout résultat insolite. Ce pourrait être quelque chose dont l'intérêt biologique est considérable, que l'on n'aurait pas remarqué autrement.

b) L'allure générale ou la forme de la relation dose-effet deviendront manifestes, ce qui pourrait prévenir l'ajustement forcé, aux données, d'un modèle mathématique non convenable.

c) Habituellement, le graphique permet d'estimer grossièrement la valeur du paramètre de toxicité. Si le paramètre estimé en vertu de l'analyse mathématique ne concorde pas suffisamment, il faut chercher la cause de l'écart. Parfois, cela pourrait aider à ne pas communiquer un résultat qui renfermait une erreur inopinée de calcul ou à ne pas effectuer un transfert erroné de données [45].

On pourrait également représenter les données brutes sur un graphique produit par l'ordinateur ou vice versa.

6.3.2 Choix de la méthode

La régression linéaire ou non linéaire est la méthode de choix pour les essais quantitatifs de toxicité sublétale effectués dans les laboratoires canadiens d'écotoxicologie. Nous donnons des conseils sur ces méthodes (§ 6.5.8), et, désormais, Environnement Canada exige la régression comme premier choix pour les essais de croissance et de reproduction des organismes vivant dans le sol (EC, 2004a, b et 2007).

La méthode la plus utilisée par le passé — et le choix le plus facile — a été le lissage et l'interpolation (v. le § 6.4). Ses défauts notables, connus des chercheurs canadiens depuis de nombreuses années, ont été décrits à la réunion de Québec du Comité consultatif de la statistique (Miller *et al.*, 1993 ; v. le § 6.4.1). Les participants y ont recommandé l'emploi de la régression comme méthode de rechange et ils ont exprimé le besoin de conseils pour la sélection du modèle approprié. Comme nous l'avons mentionné, ces conseils sont maintenant disponibles.

6.4 Lissage et interpolation

Repères

- *Pratique courante un peu partout en Amérique du Nord, cette méthode devrait progressivement cesser d'être utilisée au Canada pour faire place aux méthodes de régression. C'est une méthode commode, parce que la seule hypothèse concernant les résultats est que l'effet augmente avec la concentration.*
- *On estime la CI p par interpolation entre deux points adjacents de données, ce qui est moins avantageux que la régression, qui utilise toutes les données.*

45. Les bons programmes informatiques comprennent souvent un sous-programme utile de traçage des résultats, mais qui, comme nous l'avons fait remarquer dans le § 4.2.2, ne saurait remplacer un graphique tracé à la main. Si une erreur a eu lieu dans la saisie des données, le graphique tracé par ordinateur et les calculs arriveraient au même résultat erroné.

- *Dans un premier temps, l'analyse ajuste les données brutes en les rendant monotones, ce qui, de façon limitée, permet d'utiliser la distribution plus étendue de données.*
- *Le calcul de la CI p est suffisamment simple pour pouvoir se faire à la main, mais un gratuiciel est accessible, le programme ICPIN.*
- *À présent, le programme informatique n'utilise pas une échelle logarithmique de concentrations. Les utilisateurs canadiens du programme doivent saisir les concentrations sous forme logarithmique.*
- *Les limites de confiance ne peuvent pas être calculées par les méthodes habituelles. Plutôt, le programme informatique sert à une estimation par la méthode bootstrap. L'ordinateur rééchantillonne les mesures originelles au moins 240 fois (le minimum recommandé) pour estimer les limites de confiance.*

6.4.1 Critique générale

Cette méthode d'interpolation lancée par l'USEPA (Norberg-King, 1993) est offerte sous la forme du programme informatique ICPIN. Les estimations par interpolation linéaire souffrent de certains problèmes conceptuels (v. le texte qui suit), mais la méthode d'interpolation est polyvalente. Elle a été la façon habituelle d'obtenir une estimation ponctuelle quantitative en Amérique du Nord, faute, à l'époque, d'un progiciel de statistique commode pour effectuer la régression. Le programme ICPIN est peu connu en Europe, au moment d'écrire ces lignes (Niels Nyholm, Université technique du Danemark à Lyngby, 2001, communication personnelle)

Le Groupe consultatif sur la statistique (Miller *et al.*, 1993) a dressé la liste de certains défauts généraux de la méthode de lissage et d'interpolation, comme suit :

- Utilisation inefficace de données, puisque la méthode effectue l'interpolation seulement entre deux concentrations encadrant le paramètre de toxicité et qu'elle néglige la relation entre l'effet et la concentration, dans son ensemble (mise à part une certaine influence générale du lissage) ;
- Sensibilité à toute irrégularité ou particularité des deux concentrations utilisées ;
- Non-utilisation du logarithme de la concentration, ce qui introduit une légère majoration systématique dans le calcul de la CI p ;
- Étrécissement parfois exagéré des intervalles de confiance calculés par la méthode bootstrap.

Les trois hypothèses qui suivent sont implicitement posées dans la méthode de lissage et d'interpolation. (Parfois cette méthode est dite « sans hypothèse » parce qu'elle ne postule aucune forme particulière de courbe dose-effet, mais, néanmoins, elle pose des hypothèses.)

- Les effets doivent augmenter monotonement en passant d'une concentration à la suivante plus forte (ou, du moins, ils ne devraient pas diminuer). Si on ne satisfait pas à cette exigence, elle est imposée par les manipulations mathématiques.
- Les effets augmentent linéairement entre deux concentrations successives. (On dit parfois qu'ils suivent une *fonction linéaire par morceaux,* expression ambiguë.)
- Les effets devraient provenir d'un échantillon représentatif de données expérimentales qui est aléatoire et indépendant, hypothèse qui s'applique à la plupart des analyses statistiques.

Dans la pratique, les problèmes concernant les hypothèses requises sont rarement reconnus dans l'application de cette méthode. Pour ce qui concerne la première exigence (suite monotone), les données sont simplement ajustées, au besoin, pour rendre la suite monotone. Il n'y a pas moyen de vérifier la deuxième hypothèse (linéarité par morceaux). On ne vérifie presque jamais la troisième hypothèse (résultats aléatoires, indépendants).

L'expérimentateur a peu l'occasion de s'assurer de la fiabilité des résultats produits par la méthode. Celle-ci devrait être utilisée avec prudence, si les effets s'écartent fortement de la monotonie. La méthode est particulièrement inappropriée pour les données hormétiques (§ 10.3), comme avec certains essais avec l'algue *Pseudokirchneriella subcapitata.* Elle serait également risquée si des concentrations successives causaient des effets très petits et très grands (USEPA,

1995). Néanmoins, le lissage masquera de telles irrégularités, et la méthode est souvent utilisée lorsque les données sont irrégulières. Dans de tels cas, la prudence dicterait de soumettre les données originelles à une comparaison subjective et de tracer un graphique à la main.

6.4.2 Étapes de l'analyse

Vu que la méthode a été si largement utilisée, voici une description générale des étapes de l'estimation d'une CI 25 par lissage et interpolation. Quand un exemple est nécessaire, c'est le poids de poisson à la fin d'un essai de toxicité. L'annexe N renferme une description très détaillée de l'analyse et du programme informatique ICPIN. Les utilisateurs de la méthode feraient bien de comprendre les étapes exposées dans l'annexe N.

(1) Afin de vérifier subjectivement la qualité des données, porter sur un graphique la moyenne non ajustée de chaque groupe de poisson en fonction du logarithme de la concentration.

(2) Commencer l'interpolation linéaire par le lissage des données si le poids moyen augmente entre une concentration et la concentration supérieure suivante.

On estime la CI 25 par simple interpolation linéaire entre les deux concentrations qui l'encadrent. Les calculs, faits à la main (voir les étapes qui suivent), passent par les mêmes étapes que celles du programme informatique ICPIN. Les étapes semblent complexes, mais elles font appel, de fait, à des calculs plutôt simples.

(3) Calculer le poids correspondant au paramètre de toxicité recherché. C'est 75 % du poids moyen du poisson témoin, c'est-à-dire une réduction de 25 %.

(4) De ce poids (résultat de l'étape 3), soustraire le poids moyen à la concentration immédiatement inférieure à la CI 25. Dans une expérience de croissance, le résultat est normalement négatif.

(5) Du poids moyen à la concentration immédiatement supérieure à la CI 25, soustraire le poids à la concentration immédiatement inférieure à la CI 25. Dans une expérience de croissance, le résultat est probablement négatif.

(6) Diviser le résultat de l'étape 4 par celui de l'étape 5.

(7) Soustraire le logarithme de la concentration immédiatement inférieure à la CI 25 du logarithme de la concentration supérieure à la CI 25.

(8) Multiplier le résultat de l'étape 6 par celui de l'étape 7. Ceci représente la croissance de la concentration immédiatement inférieure à la CI 25 à cette dernière.

(9) Ajouter le résultat de l'étape 8 à la concentration logarithmique immédiatement inférieure à la CI 25. Le résultat est la CI 25 sous forme logarithmique.

La CI p ne peut pas être évaluée s'il n'existe pas une concentration expérimentale qui lui est inférieure et une autre qui lui est supérieure. Selon le cas, en effet, on pourra seulement affirmer que la CI p est inférieure (supérieure) à la plus faible (forte) concentration expérimentale.

L'ordinateur est indispensable à l'estimation des limites de confiance (v. le § 6.4.3 et l'annexe N).

6.4.3 Le programme informatique ICPIN

Le programme ICPIN roule sur les ordinateurs personnels et il est disponible dans des progiciels commerciaux ; cependant, des exemplaires gratuits sont largement accessibles (annexe N). Il est facile à utiliser, ses consignes d'emploi sont claires, et les étapes de saisie et de manipulation des données se passent d'explications.

Le programme ICPIN se charge de toutes les étapes (1 à 9) exposées dans le § 6.4.2 avant de calculer les limites de confiance. Cependant, *il faut calculer manuellement les logarithmes des concentrations d'essai et effectuer la saisie de ces logarithmes,* plutôt que des concentrations arithmétiques, contrairement à ce que l'on lit dans les consignes du programme.

Pour calculer les limites de confiance au seuil de 95 % de la CI p, l'ordinateur est indispensable. Il faut appliquer la technique dite *bootstrap,* parce que les méthodes statistiques usuelles sont inutilisables après

l'interpolation. Le programme ICPIN s'y prend en calculant une série de CI p qui auraient pu avoir été obtenues d'après les rééchantillonnages des observations originelles (v. l'annexe N). À cette fin, l'essai de toxicité doit avoir des répétitions. D'après la distribution des CI p hypothétiques, il est possible de calculer les limites de confiance de la CI p estimée.

6.5 Estimations ponctuelles par régression

Repères

- *Les techniques de régression représentent la méthode de choix pour l'estimation de la CI p. Beaucoup de publications portent sur la question, et des conseils précis en ce sens ont récemment été intégrés dans les méthodes canadiennes d'essai.*
- *La plupart des types d'effets quantitatifs sublétaux peuvent être ajustés par régression non linéaire. Nous les récapitulons en ajoutant, dans l'annexe O, des conseils détaillés point par point.*
- *Il n'existe pas de modèle unique de régression non linéaire pouvant convenir à tous les types de relations dose-effet observées. On peut ajuster la plupart de ces cas en choisissant parmi cinq modèles définis. Ensuite, on peut ajuster les données, puis estimer le paramètre de toxicité en appliquant les programmes d'un progiciel polyvalent de statistique. Le choix et l'analyse subséquente exigent des connaissances en statistique et non, simplement, l'application mécanique d'un programme informatique. Au moins un progiciel de statistique immédiatement disponible et applicable à l'écotoxicologie offre une grande sélection de modèles de régression non linéaire.*
- *La régression non linéaire peut s'ajuster aux distributions d'effets hormétiques. Un modèle proposé s'adapte à l'hormèsel Pourtant, il utilise les données obtenues du vrai témoin pour estimer le paramètre de toxicité.*
- *Environnement Canada exige désormais qu'on applique la régression dans les méthodes récentes d'essai de toxicité visant à estimer la croissance et la reproduction chez les organismes du sol.*

Le présent paragraphe donne des renseignements de base sur l'emploi de la régression, puis passe à des processus particuliers pour les méthodes de régression désormais exigées dans la plupart des essais de toxicité du sol d'Environnement Canada.

Les techniques de régression ont bénéficié de l'intérêt répandu pour l'amélioration des méthodes d'analyse des résultats des essais de toxicité sublétale. En 1995, un atelier sur les méthodes statistiques, parrainé par la SETAC (Society of Environmental Toxicology and Chemistry)-Europe a accueilli quelque deux douzaines de participants de nombreux pays (Chapman *et al.*, 1996a), et un atelier semblable a été parrainé par l'OCDE (Chapman, 1996).

Vieil outil statistique (Draper et Smith, 1981), la régression constitue probablement la meilleure méthode d'estimation des paramètres quantitatifs de toxicité sublétale. Le développement et la normalisation des méthodes de régression pour l'écotoxicologie ont été appréciables.

La régression est une description mathématique de la relation entre deux ou plusieurs variables. Dans le présent document, la *variable dépendante* est l'effet observé. Sa valeur dépend de la *variable indépendante,* la concentration, ou, peut-être, de plus d'une variable s'il existe des conditions modifiantes. Les données sont ajustées mathématiquement à un modèle choisi, puis (en toxicologie) on sélectionne, à partir du modèle, un paramètre de toxicité. Les techniques mathématiques usuelles peuvent décrire une régression pour transmettre de l'information utile. On peut prédire les effets à fortes et à faibles concentrations et l'on peut estimer des bandes de confiance. Le modèle choisi devrait se conformer aux données, même s'il ne repose sur aucune base biologique particulière ou s'il possède peu de justification théorique (Moore, 1996).

Le problème de la régression est qu'il n'existe pas un seul modèle qui s'ajuste aux diverses courbes dose-effet résultant des essais de toxicité sublétale. Un spectre de modèles est nécessaire, avec des conseils sur la façon de choisir le modèle approprié.

La *transformation* des données sur les effets est méthode utilisable pour ajuster les résultats à un modèle linéaire relativement simple. L'opération

comporte des avantages et des inconvénients (v. le § 2.9.1) et, en général, il est préférable de l'éviter.

6.5.1 Le b.a.-ba de la régression

Les exigences communes et les étapes essentielles de toute régression (dans le contexte des essais de toxicité) peuvent s'exprimer simplement comme suit :

(1) Compiler l'ensemble de données.
L'essai possède un ensemble fixe de valeurs pour la *variable indépendante* (la concentration). À chacune de ces valeurs, on fait des observations de la *variable dépendante* (l'effet).

(2) Choisir un modèle.
L'expérimentateur propose une relation entre les variables dépendantes et indépendantes. Il l'exprime sous forme d'une fonction mathématique telle qu'une droite ou une courbe logistique.

(3) Choisir une méthode d'ajustement de la relation aux données.
D'abord, on vérifie les hypothèses du modèle (par ex. normalité de données). Ensuite, on estime habituellement les paramètres du modèle en réduisant au minimum le carré des écarts entre les observations et la courbe servant de modèle. La méthode normalisée d'Environnement Canada est décrite dans le § 6.5.8.

(4) Effectuer les calculs et examiner la qualité de l'ajustement des données au modèle.

(5) Effectuer l'estimation inverse de la concentration que l'on prévoit causer le degré choisi d'effet (le paramètre de toxicité, par ex. la CI 25).

(6) Trouver les limites de confiance de ce paramètre de toxicité, également par estimation inverse.

Normalement, les calculs sont effectués par un programme informatique de régression.

6.5.2 Notions sur les modèles linéaires, non linéaires, linéaires généraux (GLM) et linéaires généralisés (GLIM)

« ... tous les modèles sont faux ; cependant, certains sont plus utiles que d'autres, et ce sont eux que nous devrions rechercher. »
(McCullagh et Nelder, 1989)

Repères

- *Dans l'expression* régression linéaire *le qualificatif* linéaire *décrit la relative simplicité de l'équation. On peut estimer les paramètres (*a, b, *etc.) en évaluant une seule formule.*
- *En régression non linéaire, les paramètres ne sont pas indépendants des autres paramètres. Il faut recourir à l'itération pour estimer les paramètres du modèle.*
- *Les modèles linéaires généraux (GLM) constituent une catégorie de modèles semblables, notamment la régression linéaire simple, l'analyse de variance, l'analyse de covariance, les mesures répétées, etc. Les modèles linéaires généralisés (GLIM) sont une catégorie élargie de l'approche employée pour les GLM. Les statisticiens s'en servent pour estimer les paramètres d'un modèle comprenant des distributions exponentielles, binomiales, logistiques, de Poisson et log-normales. La notion est tout à fait avancée et elle n'est pas encore largement utilisée en écotoxicologie.*
- *L'avantage de la régression non linéaire est de se servir de toutes les données pour une estimation ponctuelle dotée de limites de confiance pour diverses formes de courbes de la relation concentration-effet, y compris celle du phénomène de l'hormèse. La régression englobe les mesures sur le témoin. Il faut appliquer connaissances et jugement, cependant, dans le choix du modèle et l'application des méthodes statistiques.*

Rappelons aux non-mathématiciens que, pour le statisticien, les qualificatifs *linéaire* ou *non linéaire* ne décrivent pas une forme dessinée, mais qu'ils qualifient les relations des éléments avec une équation. En termes de statistique, une expression est linéaire dans son ou ses paramètres si on peut écrire une solution pour le paramètre par référence aux données uniquement et non pas à un autre paramètre. La linéarité décrit la relation entre l'effet et les *paramètres* du modèle, et non la relation entre l'effet et la ou les *variables*

indépendantes. L'exemple le plus simple de modèle linéaire est l'équation 4 d'une droite (§ 6.5.3). Les valeurs d'**α** et de **ß** peuvent être déterminées au moyen d'opérations arithmétiques, fondées sur les valeurs observées de *X* et de *Y*. En outre, une équation quadratique (équation 5) reste un modèle linéaire parce que l'on peut également estimer ses paramètres à partir des données observées.

Dans les *régressions non linéaires,* il est impossible d'estimer en une seule étape les paramètres à partir des données observées. Il faut recourir à l'itération pour résoudre les équations permettant d'estimer chaque paramètre. (v. le § 6.5.4.)

À un niveau plus complexe, l'expression « modèles linéaires » entre dans deux expressions plus longues, mais semblables au point d'entraîner la confusion et qui, pour le statisticien, possèdent des significations différentes. Les deux concernent l'analyse des résultats des essais de toxicité. Dans le premier cas, il s'agit de *modèles linéaires généraux* (*GLM*), une catégorie générale de modèles à une seule variable dépendante (§ 6.5.10). Cette catégorie comprend des modèles bien connus tels que l'analyse de variance, la régression et des modèles plus complexes tels que l'analyse de covariance (ANCOVA) et les mesures répétées. Les GLM s'appliquent seulement si les données (tels que le poids des organismes) suivent la loi normale. En vertu de cette définition rigide, les données suivant une distribution binomiale telles que les données sur la mortalité n'en feraient pas partie.

La seconde expression, *modèles linéaires généralisés* (*GLIM*), représente une catégorie encore plus grande de modèles, qui comprend notamment les GLM. Les GLIM offrent à l'expérimentateur un champ de manœuvre encore plus grand pour l'analyse des effets quantiques ou quantitatifs qui découlent de la rencontre, simple ou complexe, de variables indépendantes dans une expérience (§ 6.5.11).

6.5.3 Régression linéaire

La relation bien connue décrivant une ligne droite (équation 4) représente un modèle linéaire. (Des courbes peuvent aussi en faire partie.)

$$Y = \alpha + \beta X \qquad (4)$$

Cette formule décrit la relation entre un effet mesuré *Y,* la *variable dépendante,* et un prédicteur *X,* la *variable indépendante,* qui, dans ce cas, serait probablement le logarithme de la concentration.

Dans l'équation 4, α et β sont des paramètres. α est l'ordonnée à l'origine de la droite avec l'axe des ordonnées, c'est-à-dire la valeur de la variable dépendante (*Y*) quand la variable indépendante (*X*) est nulle. β est la pente de la régression, c'est-à-dire l'augmentation de la valeur de *Y* correspondant à chaque augmentation d'une unité de la valeur de *X.* Pour un ensemble donné, les paramètres seraient estimés par quelque méthode mathématique. Souvent, on emploie à cette fin la *méthode des moindres carrés*, qui permet d'estimer les paramètres permettant de réduire au minimum la somme des carrés des écarts des valeurs observées par rapport au modèle.

La relation pourrait être causale comme le laisse entendre le qualificatif « dépendant » ou elle pourrait n'être qu'une corrélation. Comme il n'y a que deux variables à considérer, il s'agit d'une *régression simple,* d'où l'expression *régression linéaire simple* (Zar, 1999). [Il pourrait y avoir plus d'une variable indépendante, auquel cas où une formule plus complexe décrirait la relation, comme il est précisé plus loin, sous la rubrique « Régression multiple ».]

Les essais de toxicité pourraient parfois révéler une relation entre l'effet et la concentration qui semble directement proportionnelle, du moins dans la partie centrale de la régression. Le calcul du meilleur ajustement de la droite (modèle) aux données pourrait alors se faire par les moyens classiques que l'on emploie avec les régressions linéaires, telles que la méthode des moindres carrés. En effet, les régressions simples ont servi à décrire des résultats, particulièrement pour les effets sublétaux tels que la croissance (par ex. Rowe *et al.*, 1983).

La régression linéaire est un modèle simple. Si un ensemble de données toxicologiques s'ajuste bien à la régression, on peut se servir de cette dernière pour formuler des prévisions. Pour toute valeur donnée ou choisie de *X* (par ex. le logarithme de la concentration), on peut calculer, à partir de l'équation, la valeur prévue de *Y* (disons le poids des poissons exposés à cette concentration). Il importe que les valeurs de la variable

indépendante X soient créées et mesurées sans erreur. [Comme il est décrit dans les § 6.2.4 et 9.4, le toxicologue effectue finalement une inversion pour la concentration (et ses limites de confiance) réputée devoir causer un degré choisi d'effet (par ex. réduction de 25 % des performances par rapport à ceux du témoin, la CI p.)]

Une description plus complète, plus juste et plus explicite du modèle ajouterait des indices à l'équation 4. Bien que nous les omettions dans le présent document, ils sont implicites, et les expérimentateurs devraient s'attendre à en rencontrer dans d'autres sources. Des indices seraient nécessaires si une équation représentait un ensemble d'observations dans un essai. L'indice i désignerait chacun des organismes ou chacune des mesures de l'essai, tandis que l'indice j désignerait les concentrations de toxique. L'équation 4 deviendrait donc l'équation 4a, comme suit :

$$Y_{ij} = \alpha + \beta X_j \qquad (4a)$$

Comme les points de données seraient dispersés de part et d'autre de la droite ajustée, on ajoute à l'équation un terme d'erreur (ε_{ij} ou e_{ij}). Le terme e représente la variabilité aléatoire d'une mesure individuelle i à la j-ième concentration. La *régression linéaire* complète est l'équation 4b.

$$Y_{ij} = \alpha + \beta X_j = \varepsilon_{ij} \qquad (4b)$$

Régressions multiples. — Celles-ci font partie de la catégorie des régressions linéaires. L'expression signifie que la variable dépendante est sous l'emprise de deux ou de plusieurs variables indépendantes (X_1 et X_2 dans l'équation 5). Par exemple, la toxicité d'un métal pourrait dépendre non seulement de la concentration du métal, mais aussi de la température du milieu. L'équation 5 pourrait représenter une régression à quatre paramètres : α, β_1, β_2 et la variance (σ).

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2^2 \qquad (5)$$

L'équation 5 correspond à plusieurs catégories. On peut la qualifier de *régression multiple* parce qu'elle comprend plusieurs termes. C'est aussi une *fonction quadratique,* en raison de l'ajout, à la fin, du terme dit quadratique. Les statisticiens font remarquer que, appliqué aux données sur la toxicité, ce modèle devrait se borner à décrire un effet *local.* Théoriquement, l'équation quadratique ne convient pas, parce qu'elle prédit une diminution de l'effet à une forte concentration, au lieu de la situation habituelle de la relation dose-effet. Comme nous l'avons mentionné, elle peut cependant être utile pour décrire des effets locaux à l'intérieur d'une étendue limitée de concentrations.

6.5.4 Aspects généraux des régressions non linéaires

La relation linéaire (§ 6.5.2) est une relation relativement simple, souvent insuffisante pour décrire une relation complexe de l'évolution de l'effet avec la concentration. L'expérimentateur devrait choisir un *modèle* (c'est-à-dire une fonction mathématique) plus complexe afin d'ajuster les données sur la toxicité ; la forme de la relation dose-effet pourrait bien mener à l'adoption d'un modèle de régression non linéaire. Deux paramètres ou plus du modèle pourraient être des fonctions l'un de l'autre, de façon multiplicative, comme dans la croissance exponentielle, montrée par l'équation 6 (Zar, 1999). Manifestement, l'estimation des paramètres d'une telle équation sera plus compliquée que celle des paramètres d'une régression linéaire.

$$Y = \alpha \beta^X \qquad (6)$$

Souvent, une fonction décrivant une forme sigmoïde conviendrait en écotoxicologie. Deux modèles non linéaires se sont souvent révélés convenir : le modèle logit et l'équation de Weibull. Le modèle logit est symétrique tandis que le modèle de Weibull est asymétrique (v. les § 4.5.1, 4.5.2 et l'annexe J). D'autres modèles utiles sont exposés en détail dans le § 6.5.8.

Une fois la fonction (le modèle) spécifiée, on trouve les « meilleures » estimation de ses paramètres par la technique du maximum de vraisemblance ou celle des moindres carrés. Comme nous l'avons mentionné, il faut procéder par itération pour résoudre les équations estimant chaque paramètre.

L'itération appliquée aux régressions non linéaires pourrait se décrire simplement comme des « suppositions » initiales des valeurs des paramètres du modèle, faites par l'expérimentateur ou le programme utilisé. Au moyen d'itérations successives, ces valeurs initiales sont modifiées (majorées ou diminuées) par le programme, pour mieux les approcher d'un ajustement des données observées. Autrement dit, le programme cherche une valeur optimale pour chaque paramètre. On peut visualiser le modèle comme un groupe de petites collines représentant les divers paramètres, chaque paramètre possédant une valeur optimale au sommet d'une colline. Le programme peut déterminer au moment de chaque itération, la pente de la colline localement et, de là, la bonne direction vers où se diriger pour la prochaine itération afin de se rapprocher de la valeur optimale du paramètre (le « sommet de la colline »). Quand les estimations de tous les paramètres restent essentiellement constantes d'une itération à l'autre, l'opération a *convergé* vers une solution finale, c'est-à-dire qu'elle est parvenue aux meilleures estimations des paramètres du modèle pour l'ensemble particulier de données.

L'OCDE (OECD, 2004) fait remarquer qu'il peut être important de faire des « suppositions » initiales réalistes des valeurs des paramètres. Les estimations finales pourraient dépendre de ce choix initial, parce qu'il pourrait se trouver plusieurs maximums ou optimums locaux d'un paramètre donné. Cela pourrait se représenter sous la forme de plusieurs petites proéminences dispersées sur les pentes d'une grande colline. Comme le programme, en une itération quelconque, peut n'apprécier que la pente se trouvant dans le voisinage immédiat et non la pente de la colline, il pourrait gagner le sommet d'une proéminence et y rester, en vertu d'une « convergence » indésirable. D'où l'importance d'un point de départ réaliste, près du sommet principal.

Une méthode pour parvenir aux estimations finales des variables d'une équation consiste à utiliser la technique des moindres carrés. Il a été fait allusion à l'itération de la méthode des moindres carrés pour la résolution de la régression probit de données quantiques (§ 4.5.3). Dans la méthode des moindres carrés, les valeurs prévues et observées de la variable dépendante (l'effet toxique) sont comparées à des concentrations données de la variable indépendante (le logarithme de la concentration). La différence entre la valeur prévue et la valeur observée s'appelle *résidu,* et mieux la droite est ajusté, plus les résidus diminuent. On élève les résidus au carré et on les additionne. C'est cette « somme des carrés » que l'on prend comme mesure de l'ajustement. Manifestement, la somme minimale des carrés correspond au meilleur ajustement, d'où l'expression « méthode des moindres carrés ».

La solution trouvée par la méthode des moindres carrés aux paramètres d'une équation équivaut souvent à la *solution de maximum de vraisemblance,* méthode mathématique plus recherchée, plus complexe et plus mystérieuse.

Comme nous l'avons écrit dans les § 6.5.7 et 6.5.8, la régression non linéaire exige un certain jugement et une connaissance des techniques mathématiques. On utilise souvent un logiciel général de statistique, bien qu'il existe au moins un progiciel de statistique conçu spécifiquement pour l'écotoxicologie (CETIS), qui offre un large choix de modèles. Les méthodes générales de régression existent depuis un certain temps dans les manuels et les progiciels classiques de statistique, mais il a souvent fallu du temps aux toxicologues pour développer leurs propres compétences (Moore, 1996). Les techniques utiles en toxicologie ont été décrites par Newman (1995).

On trouve des conseils sur la régression non linéaire dans les manuels tels que ceux de Bates et Watts (1988). L'expérimentateur qui commence à utiliser la régression non linéaire bénéficierait des conseils d'un statisticien expérimenté (§ 2.1). L'expérimentateur naïf pourrait obtenir des résultats erronés en ne satisfaisant pas aux hypothèses techniques, en choisissant un modèle inapproprié, etc. De plus amples renseignements sont donnés dans les § 6.5.7 à 6.5.9.

Avantages de la régression non linéaire. — Pour l'analyse des données sur la toxicité, la régression est beaucoup plus défendable que le lissage et l'interpolation ou que le test d'hypothèse. Les données expérimentales dicteront le type de régression à appliquer. Si la régression linéaire est ajustée aux données, on devrait l'utiliser ; sinon, on privilégie le modèle non linéaire. On peut énumérer ci-dessous certains des avantages généraux de la régression et les avantages particuliers de la régression non linéaire :

- On utilise tous les résultats de l'essai ;
- On obtient une estimation ponctuelle, la CI p ;
- On obtient les limites de confiance de la CI p ;
- On peut utiliser n'importe quelle valeur de *p,* par ex. la CI 25 ; on peut prendre en charge diverses formes de courbes concentration-effet ;
- On tient compte, dans la régression ajustée, des résultats obtenus chez le témoin ;
- On peut prendre en charge l'hormèse sans compromettre l'effet observé chez les témoins.

Le principal inconvénient est qu'il ne peut pas y avoir de programme informatique simple, de style « boîte noire », conçu pour la toxicologie. L'expérimentateur doit utiliser connaissances et jugement dans la sélection du modèle et l'application des méthodes statistiques.

6.5.5 Choix d'un modèle de régression

Repères

- *Il est prudent de choisir un modèle adéquat, mais aussi simple que possible. Il est souhaitable de respecter le principe de parcimonie des paramètres — chaque paramètre ajouté au modèle lui fait perdre un degré de liberté.*
- *Une façon de conserver au modèle sa simplicité est d'éliminer les paramètres corrélés à un autre paramètre déjà modélisé.*
- *Le modèle pourrait ne pas être ajusté parce qu'on l'a mal choisi, qu'on en a choisi un excessivement compliqué, que des observations sont aberrantes ou que des erreurs sont survenues au codage. Parfois, les données originelles pourraient ne pas couvrir le haut ou le bas du domaine du modèle.*
- *Le graphique des résidus en fonction de la valeur prévue permet l'évaluation visuelle de l'ajustement du modèle, et on a toujours besoin d'une vérification visuelle des données. Dans le cas de la régression linéaire, le coefficient de détermination (la valeur* R^2*) peut servir à évaluer l'ajustement.*

En choisissant un modèle, l'expérimentateur doit prendre en considération ses propres priorités, de même que les aspects techniques. Le « meilleur » modèle pour une personne pourrait être celui qui renferme le moins d'erreurs de prévision, tandis qu'une autre personne pourrait insister sur le respect du principe de parcimonie ou un autre, encore, préférerait le modèle qui éclairerait le plus les mécanismes biologiques. Certains aspects intervenant dans le choix du modèle sont mentionnés dans le texte qui suit.

Il est prudent d'adopter des modèles relativement simples mais adéquats et d'éviter les modèles excessivement complexes. Certes, on pourrait ajuster une équation polynomiale dotée d'un nombre suffisant de termes à presque toute forme inhabituelle d'effet, mais l'ajout de paramètres supplémentaires entraîne des pénalités telles que la perte de degrés de liberté et l'élargissement de l'intervalle de confiance. Dans leur texte sur des modèles linéaires généralisés, McCullagh et Nelder (1989) déconseillent d'employer beaucoup de paramètres pour obtenir un ajustement précis aux données.

> « Ce faisant, cependant, nous n'avons obtenu aucune réduction de la complexité ... la simplicité, représentée par la parcimonie des paramètres, est aussi une qualité de n'importe quel modèle ; nous n'incluons pas de paramètres dont nous n'avons pas besoin. Non seulement un modèle parcimonieux permet-il au chercheur ... de penser à ses données, mais un modèle qui est en grande partie juste donne de meilleures prévisions qu'un modèle comprenant des paramètres supplémentaires inutiles. »

Dans cette citation, « inutiles » pourrait être interprété comme qualifiant un paramètre statistiquement non significatif. Un autre spécialiste attire l'attention sur la possibilité qu'une interprétation biologique obscure découle d'un modèle complexe d'équation à quatre paramètres :

> « un ajustement pourrait sembler génial, mais comment utiliser les résultats quand les informaticiens sont inaccessibles ? » (Nyholm, 2001).

Un exemple de complexité inutile serait des observations fortement corrélées (par ex. longueur et poids des organismes), et l'expérimentateur devrait se méfier de l'emploi de paramètres pour chacune dans un modèle de régression. Cela peut mener à un problème « de multicolinéarité » et à des messages d'erreur ou à un manque d'ajustement. Les progiciels de statistique

produisent d'habitude une matrice de corrélations pour les paramètres et on devrait l'examiner ; les fortes corrélations pourraient indiquer qu'une des variables d'une paire de variables pourrait être omise.

Dans les régressions multiples, il est possible de vérifier — et cela est fortement recommandé — si toutes les variables sont nécessaires. À cette fin, il est préférable d'effectuer une série d'ajustements, avec et sans les paramètres auxquels on s'intéresse et de comparer les résultats (v. la rubrique « Explication de la variabilité de la régression » dans le § 6.5.6). Une autre méthode, mentionnée dans certains manuels, mais que nous ne recommandons pas, consiste à vérifier chaque paramètre au moyen du test *t* (parfois fourni avec le progiciel de statistique). L'hypothèse nulle serait que le paramètre égale zéro et que si le test *t* ne le réfute pas, le paramètre est supprimé de la régression.

La pondération pourrait être nécessaire, comme l'expliquent Nyholm *et al.* (1992) :

> « si la variance des points de données est constante (erreur absolue constante), on peut effectuer une régression non linéaire, directement, sans pondération,... sinon il faut une pondération statistique appropriée. Les coefficients de pondération devraient être inversement proportionnels à la variance des points de données ... ».

Cela exige des répétitions et la vérification de l'équivariance, comme il est décrit dans le § 6.5.8.

6.5.6 Adéquation et ajustement

Manque d'ajustement. — Le modèle pourrait « ne pas réussir à converger » auquel cas il n'y aurait pas d'ajustement ni d'estimation des paramètres. La multicolinéarité peut expliquer la non-convergence (v. le § 6.5.5). Parfois, on pourrait ne pas obtenir d'estimations satisfaisantes des paramètres, même après convergence satisfaisante. On peut énumérer pour cela quelques raisons possibles :

- **Mauvais choix de modèle.** — On ne peut pas s'attendre qu'un modèle qui ne convient pas puisse être ajusté.

- **Observations aberrantes.** — Même une observation aberrante pourrait empêcher la convergence. Cette observation ne doit pas être arbitrairement retranchée du processus de modélisation. Il faut plutôt en tenir compte objectivement par des méthodes telles que celles qui sont mentionnées dans les § 6.5.8 et 10.2.

- **Erreurs de codage.** — Des inexactitudes ou des erreurs dans le codage (v. le glossaire) peuvent donner des résultats absurdes.

- **Étendue des données.** — Les méthodes pourraient être satisfaisantes, mais les données originelles pourraient être déficientes. Les valeurs pourraient ne pas couvrir le domaine supérieur ou inférieur du modèle. Les données quantiques devraient couvrir toute l'étendue des effets, de l'absence d'effet à l'effet total. Les données quantitatives devraient être représentées dans chaque branche de la courbe du modèle. C'est une carence relativement fréquente, dont il a été question dans le § 2.2. On peut y remédier par une recherche de la gamme de concentrations à utiliser.

- **Modèle trop compliqué.** — (Par ex. multicolinéarité, § 6.5.5). On devrait adopter un modèle plus simple si les observations ne couvrent pas une partie de la distribution voulue.

Quand les paramètres ont bien été estimés, l'expérimentateur doit décider si le modèle décrit convenablement la variabilité. La plupart des progiciels de statistique offrent un test F ; si ce test donne pour *p* une valeur inférieure à 0,05, on peut conclure que le modèle de régression décrit une proportion significative des données, au niveau de confiance de 95 %. L'autre évaluation devrait se poursuivre de la façon décrite ci-dessous.

Explication de la variabilité dans la régression. — Le *graphique des résidus* en fonction des valeurs prévues permet l'évaluation visuelle de la qualité de l'ajustement du modèle (v. la notion de *résidu* dans le glossaire.). On peut, de la sorte, révéler certains problèmes. Une série de résidus supérieurs ou inférieurs aux valeurs prévues pourrait traduire un ajustement insatisfaisant ou une corrélation insatisfaisante des observations. La distorsion des résidus en forme de V, dans le graphique, traduit une hétérogénéité de la variance. Si la dispersion est divergente, cela est le signe d'un modèle qui ne convient pas (v. le § 6.5.8 et l'annexe O).

D'autres évaluations fondées sur le bon sens devraient suivre le tracé du graphique. L'intervalle des concentrations testées était-il suffisamment étendu pour révéler l'étendue des effets ? Le tracé de la régression ajustée représente-t-il de façon convenable les observations réelles. La forme du modèle est-elle ajustée aux mécanismes que l'on pense gouverner l'effet ? Des observations aberrantes ont-elles exagérément influé sur l'ajustement ? Si, à ces questions, les réponses négatives sont plus nombreuses, l'expérimentateur serait bien avisé de consulter un statisticien.

Dans le cas de la régression linéaire, le *coefficient de détermination* ou R^2 (« la valeur R^2 ») est la somme des carrés (SC) expliqués par l'ajustement du modèle ($SC_{régression}$) divisée par la somme totale des carrés (SC_{totale}) par rapport à la moyenne. Les valeurs sont souvent exprimées en pourcentage et pourraient, théoriquement, aller de 0 (modèle n'expliquant rien) à 100 % (ajustement parfait du modèle). Ce taux de 100 % ne sera pas observé, et des résultats très élevés ne sont pas nécessairement souhaitables. De tels résultats portent à croire en un modèle complexe, aux nombreux paramètres et aux inconvénients associés (v. le § 6.5.5). Le coefficient de détermination ne peut pas s'appliquer à des modèles non linéaires.

L'OCDE (OECD, 2004) déconseille l'application aveugle d'un test statistique de la qualité de l'ajustement d'une façon stricte et absolue (c'est-à-dire soit le modèle est ajusté, soit il ne l'est pas). On lit dans son guide qu'une « vérification visuelle des données est toujours nécessaire et peut prévaloir sur un test d'ajustement. » Ce conseil vise à encourager l'expérimentateur à vérifier que les données communiquent suffisamment de renseignements pour confiner le modèle. Par exemple, si on avait disposé de données supplémentaires sur les doses intermédiaires, cela pourrait-il avoir modifié la forme de la relation ? L'OCDE fait aussi observer que les données correspondant à un petit nombre de traitements peuvent plus facilement réussir le test d'ajustement. En revanche, un bon ensemble de données comportant un seul traitement ou effet aberrant pourrait entraîner le rejet d'un modèle qui, par ailleurs, était parfaitement ajusté aux données.

On peut évaluer l'ajustement par d'autres moyens. L'analyse de variance peut résumer un modèle de régression, et le test F vérifie globalement l'hypothèse nulle de l'ajustement convenable. Une autre forme de R^2 n'utiliserait que le dénominateur pour décrire l'erreur résiduelle. Une petite valeur est souhaitable, mais, encore une fois, le paramétrage à outrance peut être une cause de faible erreur. Mallows fournit une version supérieure de R^2 (1973), dont C_p pénalise les modèles paramétrés à outrance. Des mesures semblables, qui devraient être reconnues comme supérieures si on en rencontre, sont le critère bayésien d'information et le critère d'information d'Akaike.

6.5.7 Exemples récents de régressions non linéaires

Repères

- *Un groupe d'auteurs canadiens a mis au point des méthodes ayant inspiré l'approche normalisée d'Environnement Canada consistant à utiliser la régression pour les estimations ponctuelles de paramètres quantitatifs de toxicité sublétale,*
- *Ces auteurs ont appliqué des modèles de régression linéaire et non linéaire qui étaient offerts dans un progiciel usuel de statistique (SYSTAT) aux résultats de leurs essais de toxicité du sol pour les plantes.*
- *Ils ont constaté que l'on pouvait ajuster de façon satisfaisante à la plupart des ensembles de résultats l'un des cinq modèles suivants : linéaire, logistique, logistico-hormétique, exponentiel et de Gompertz.*

Stephenson *et al.* (2000) ont présenté des illustrations claires de l'ajustement de régressions non linéaires aux données quantitatives sur la toxicité sublétale, tandis que Koper (1999) a expliqué la même recherche. Ces chercheurs ont obtenu des estimations utiles de la toxicité sublétale de sols contaminés pour plusieurs espèces de végétaux. Leurs méthodes ont été développées plus avant par Environnement Canada en tant que méthodes exigées dans les nouveaux essais de sol (EC, 2004a, b et 2007 ; v. le § 6.5.8).

Stephenson *et al.* (2000) ont illustré la forme générale et expliqué les équations de trois modèles : logistique,

hormético-logistique et exponentiel. D'autres modèles utiles ont été ajoutés à leur progiciel de modèles de régression non linéaire (Koper, 1999). L'un était le modèle sigmoïde de Gompertz, un autre une équation linéaire ordinaire. On a ajouté un paramètre au modèle exponentiel pour permettre à l'asymptote d'être une valeur non nulle.

Les techniques de régression non linéaire et les difficultés que celle-ci pose sont décrites brièvement par Stephenson *et al.* (2000), qui fournissent un organigramme semblable à celui de la fig. 16 pour aider à s'y retrouver dans la sélection du modèle le plus approprié. L'expérimentateur aura dû faire une estimation initiale de chaque paramètre du modèle. (Il faut faire des estimations initiales réalistes, sinon le programme de statistique pourrait choisir un paramètre de toxicité anormal : v. l'annexe O.) Les paramètres de l'équation ajustée ont ensuite été estimés de calculs itératifs. Stephenson *et al.* (2000) ont fait observer qu'un nombre excessif de paramètres pourrait empêcher les estimations. Une stratégie fructueuse a consisté à utiliser le modèle approprié le plus simple (§ 6.5.5), un nombre suffisant de répétitions et jusqu'à 12 traitements. La nécessité de variances égales pour les traitements posait également problème, parce que des variances inégales pourraient mener à une estimation gonflée de l'erreur type et des limites de confiance. C'est pourquoi on a pondéré les observations à l'aide de l'inverse de la variance pour les observations correspondant à chaque traitement (v. le § 2.6). De bonnes estimations de la variance étaient nécessaires à cette fin, parfois au moins 9 répétitions par concentration. Koper (1999) a recommandé que si la pondération était nécessaire, il faudrait effectuer des calculs pour les distributions pondérées et non pondérées, puis on pourrait comparer les résultats et la distribution des résidus.

Koper (1999) a fait remarquer que, grâce à l'ordinateur, la régression non linéaire était devenue réalisable en routine dans les laboratoires. Les modèles ont été reparamétrés pour une estimation automatique de la CI p et de ses limites de confiance (v. le § 6.5.12, sur le reparamétrage). Le reparamétrage a été inspiré par les méthodes de Van Ewijk et Hoekstra (1993) ainsi que d'Hoekstra et Van Ewijk (1993). Les analyses ont employé le progiciel de statistique SYSTAT 7.0.1.

Les problèmes d'ajustement pourraient être imputables à la colinéarité, qui survient quand les paramètres sont fortement corrélés ou quand une valeur près de zéro dans le dénominateur d'une matrice a été inversée à la faveur des calculs. D'autres causes d'éventuelles difficultés statistiques ont été la convergence, le choix d'un algorithme de maximisation, les maximums locaux par opposition aux maximums globaux et la comparaison de modèles emboîtés et non emboîtés.

Les méthodes de Stephenson *et al.* (2000) étaient assorties de certaines exigences. Les données devaient encadrer la CI p (ce qui serait utile ou essentiel à d'autres méthodes). Au moins 10 ou 12 traitements étaient recommandés, pour montrer la forme de la relation et permettre le choix du modèle. Le nombre élevé de traitements a aussi contribué à la réussite des calculs informatiques. Le nombre de répétitions par traitement pourrait être de deux, bien que ces chercheurs en aient employé jusqu'à six. Il n'était pas nécessaire que le nombre de répétitions soit identique à chaque concentration.

Les lecteurs qui voudraient imiter ce travail devraient être conscients que Stephenson *et al.* (2000) n'ont pas utilisé les logarithmes de la concentration. Or, il faudrait les utiliser dans les diagrammes de dispersion et les calculs, comme dans le mode opératoire normalisé d'Environnement Canada (§ 6.5.8). Comme nous l'avons expliqué dans le § 2.3, c'est une question de méthode scientifique et non pas simplement une question de méthode statistique ou de savoir si le modèle est capable de traiter les valeurs arithmétiques des concentrations.

6.5.8 La méthode de régression d'Environnement Canada

Repères

- *Dans ses nouveaux essais de toxicité d'un sol, Environnement Canada exige, comme premier choix pour l'estimation de la CI p, la régression linéaire ou non linéaire. Des méthodes particulières sont exposées pour SYSTAT ou d'autres progiciels de statistique.*
- *L'expérimentateur a le choix entre cinq modèles : linéaire, logistique, exponentiel, logistique adapté*

à l'hormèse (hormético-logistique), de Gompertz. Les modèles ont été reparamétrés pour permettre l'estimation directe de la CI p et de ses limites de confiance.

- *Avant d'effectuer l'estimation, il faut satisfaire aux hypothèses de normalité et d'homoscédasticité des résidus.*
- *Si les méthodes de régression ne sont pas couronnées de succès, il faut estimer la CI p par interpolation, à l'aide du programme ICPIN.*

Les nouvelles méthodes d'essai biologique publiées par Environnement Canada, dans lesquelles on emploie des vers de terre, des végétaux et des collemboles (EC, 2004a, b et 2007) exigent, comme méthode préliminaire d'analyse des données quantitatives sur la toxicité sublétale, l'application de la régression linéaire et non linéaire. C'est seulement si les résultats ne se prêtent pas à la régression que l'expérimentateur est autorisé à recourir à des méthodes moins souhaitables d'analyse.

Après un essai à plusieurs concentrations, il faut calculer la CI p et ses limites de confiance au seuil de 95 % à l'aide d'un ou de plusieurs modèles de régression linéaire et non linéaire proposés par Stephenson *et al.* (2000). Les modèles ont été reparamétrés à l'aide des techniques de van Ewijk et Hoekstra (1993), pour donner automatiquement la CI p et ses limites de confiance au seuil de 95 % pour toute valeur spécifiée de *p* (par ex. la CI 25 ou la CI 50). Les modèles comprennent un modèle linéaire et les quatre modèles suivants de régression non linéaire : exponentiel, de Gompertz, logistique et hormético-logistique [46]. Des consignes sont données dans l'annexe O pour appliquer la régression linéaire et non linéaire à l'aide de la version 11.0 du programme de statistique SYSTAT [47]. Cependant, on peut utiliser tout logiciel de statistique capable d'effectuer une régression linéaire et non linéaire (voir, à la fin du présent paragraphe, des observations sur d'autres logiciels de statistique).

Ci-dessous, suivent les descriptions des cinq modèles, avec de plus amples renseignements dans l'annexe O. Le modèle exponentiel est une version générale, tandis que la version codée de l'annexe O comporte des modifications particulières.

Modèle exponentiel

$$Y = a \times (1 - p)^{(C \div CIp)}$$

Où :

- Y = la variable dépendante (par ex. le nombre de jeunes, la longueur des racines ou des pousses, la masse sèche) ;
- a = l'ordonnée à l'origine (c'est-à-dire la réaction des organismes témoins) ;
- p = la valeur spécifiée de p, par ex. 0,25 si l'inhibition est de 25 %) ;
- C = le logarithme de la concentration d'essai ;
- $CI\,p$ = la CI p estimée pour l'ensemble des données.

Modèle de Gompertz

$$Y = t \times \exp[\log(1 - p) \times (C \div CI\,p)^b]$$

Où :

- Y = la variable dépendante (par ex. le nombre de jeunes, la longueur des racines ou des pousses, la masse sèche) ;
- t = l'ordonnée à l'origine (c'est-à-dire la réaction des organismes témoins) ;
- exp = l'exposant de la base du logarithme népérien ;
- p = la valeur spécifiée de p, par ex. 0,25 si l'inhibition est de 25 %) ;
- C = le logarithme de la concentration d'essai ;

46. On pourrait observer une réaction hormétique (stimulation à faible dose) dans les observations d'un effet sublétal correspondant à la ou aux concentrations les plus faibles, c'est-à-dire à une augmentation des performances à ces concentrations par rapport à celles du témoin. Par exemple, la progéniture engendrée à de faibles concentrations pourrait être plus nombreuse que chez le témoin ou les individus pourraient être plus lourds que les témoins. Cette réaction est un véritable phénomène biologique et non le résultat d'une faille dans l'essai. On devrait analyser ces données au moyen du modèle hormétique. Les effets hormétiques sont pris en compte dans la régression, mais ils ne biaisent pas l'estimation de la CI p. La CI 25 estimée continuerait de correspondre à une réduction de 25 % des performances par rapport au témoin.

47. On peut acheter la version la plus récente (à partir de 11.0) de SYSTAT[MD] auprès de SYSTAT Software, Inc., 501, Canal Boulevard, Suite C, Point Richmond, Calif., 94804-2028, États-Unis, tél : 800-797-7401 ; www.systat.com/products/Systat/.

$CI\,p$ = la CI p estimée pour l'ensemble des données ;
b = un paramètre d'échelle, estimé entre 1 et 4, qui définit la forme de l'équation.

Modèle hormétique

$$Y = t \times [1 + (h \times C)] \div \{ 1 + [(p + (h \times C)) \div (1 - p)] \times (C \div CI\,p)^b \}$$

Où :
Y = la variable dépendante (par ex. le nombre de jeunes, la longueur des racines ou des pousses, la masse sèche) ;
t = l'ordonnée à l'origine (c'est-à-dire la réaction des organismes témoins) ;
h = un descripteur de l'effet hormétique, estimé petit, habituellement entre 0,1 et 1) ;
C = le logarithme de la concentration d'essai ;
p = la valeur spécifiée de p, par ex. 0,25 si l'inhibition est de 25 %) ;
$CI\,p$ = la CI p estimée pour l'ensemble des données ;
b = un paramètre d'échelle, estimé entre 1 et 4, qui définit la forme de l'équation.

Modèle linéaire

$$Y = [(-\,b \times p) \div CI\,p] \times C + b$$

Où :
Y = la variable dépendante (par ex. le nombre de jeunes, la longueur des racines ou des pousses, la masse sèche) ;
b = l'ordonnée à l'origine (c'est-à-dire la réaction des organismes témoins) ;
p = la valeur spécifiée de p, par ex. 0,25 si l'inhibition est de 25 %) ;
$CI\,p$ = la CI p estimée pour l'ensemble des données ;
C = le logarithme de la concentration d'essai.

Modèle logistique

$$Y = t \div \{1 + [\,p \div (1 - p)] \times (C \div CI\,p)^b \}$$

Où :
Y = la variable dépendante (par ex. le nombre de jeunes, la longueur des racines ou des pousses, la masse sèche) ;
t = l'ordonnée à l'origine (c'est-à-dire la réaction des organismes témoins) ;
p = la valeur spécifiée de p, par ex. 0,25 si l'inhibition est de 25 %) ;
C = le logarithme de la concentration d'essai ;
$CI\,p$ = la CI p estimée pour l'ensemble des données ;
b = un paramètre d'échelle, estimé entre 1 et 4, qui définit la forme de l'équation.

Le processus général de sélection du modèle de régression le plus approprié et de l'analyse statistique ultérieure des données quantitatives sur la toxicité est exposé dans la fig. 16. Il débute par l'examen d'un diagramme de dispersion (nuage de points) ou d'un graphique linéaire représentant les données expérimentales, pour déterminer la forme de la courbe concentration-réponse, que l'on compare ensuite à celle des modèles disponibles, pour choisir, en vue d'un examen approfondi, le ou les modèles correspondant le mieux aux données (v. la fig. O.1 de l'annexe O pour des exemples des cinq modèles).

Une fois le ou les modèles choisis pour examen approfondi, on évalue les hypothèses de normalité et l'homoscédasticité des résidus. Si la régression d'un ou de plusieurs des modèles examinés satisfait aux hypothèses, on examine les données (et la régression) pour y déceler d'éventuelles observations aberrantes. Le cas échéant, on devrait examiner les procès-verbaux de l'essai et les conditions expérimentales pour y dépister des traces d'erreur humaine. Ensuite, on devrait effectuer l'analyse avec et sans l'observation ou les observations aberrantes, afin de déterminer leur effet sur la régression. Il faut prendre une décision sur la suppression ou non des observations aberrantes de l'analyse finale, compte tenu de la variation biologique naturelle et d'autres causes biologiques de l'anomalie apparente. Nous donnons des conseils supplémentaires sur la présence d'observations aberrantes et inhabituelles dans le § O.2.4 de l'annexe O ainsi que dans le § 10.2. Il est également conseillé d'obtenir l'avis d'un statisticien qui connaît bien la conduite à tenir avec les observations aberrantes.

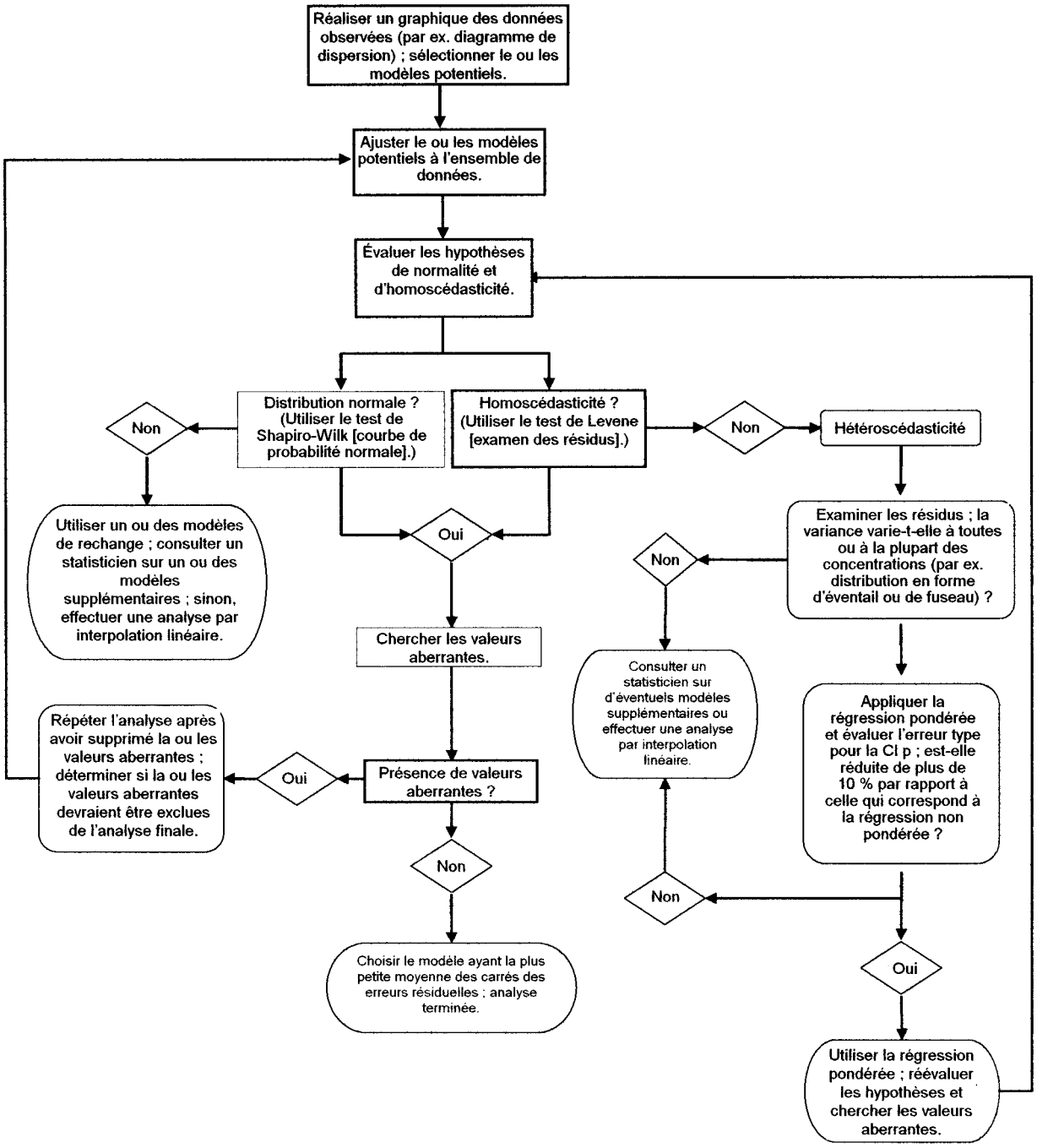

Réaliser un graphique des données observées (par ex. diagramme de dispersion) ; sélectionner le ou les modèles potentiels.
Ajuster le ou les modèles potentiels à l'ensemble de données.
Évaluer les hypothèses de normalité et d'homoscédasticité.
Distribution normale ? (Utiliser le test de Shapiro-Wilk [courbe de probabilité normale].)
Homoscédasticité ? (Utiliser le test de Levene [examen des résidus].)
Non
Non
Hétéroscédasticité
Utiliser un ou des modèles de rechange ; consulter un statisticien sur un ou des modèles supplémentaires ; sinon, effectuer une analyse par interpolation linéaire.
Oui
Examiner les résidus ; la variance varie-t-elle à toutes ou à la plupart des concentrations (par ex. distribution en forme d'éventail ou de fuseau) ?
Non
Chercher les valeurs aberrantes.
Consulter un statisticien sur d'éventuels modèles supplémentaires ou effectuer une analyse par interpolation linéaire.
Appliquer la régression pondérée et évaluer l'erreur type pour la CI p ; est-elle réduite de plus de 10 % par rapport à celle qui correspond à la régression non pondérée ?
Répéter l'analyse après avoir supprimé la ou les valeurs aberrantes ; déterminer si la ou les valeurs aberrantes devraient être exclues de l'analyse finale.
Oui
Présence de valeurs aberrantes ?
Non
Non
Choisir le modèle ayant la plus petite moyenne des carrés des erreurs résiduelles ; analyse terminée.
Oui
Utiliser la régression pondérée ; réévaluer les hypothèses et chercher les valeurs aberrantes.

Figure 16. — Organigramme général de la sélection du modèle le plus approprié et de l'analyse statistique des données sur la toxicité quantitative (adapté et modifié de Stephenson *et al.,* 2000).

Si aucune observation n'est aberrante ou si on n'en supprime aucune de l'analyse finale, on retient le modèle présentant la plus petite moyenne des carrés des erreurs résiduelles.

On devrait évaluer la normalité à l'aide du *test de Shapiro-Wilk*, décrit dans les § P.2.1 et P.2.2 de l'annexe P. On peut utiliser, pendant la régression, le tracé de probabilité normale des résidus, mais cela n'est pas recommandé comme test unique de la normalité, parce que la détection d'une distribution « normale » ou « non normale » dépendrait de l'évaluation subjective de l'utilisateur. Si les données ne pas suivent la loi normale, il est conseillé d'essayer un autre modèle, de consulter un statisticien pour obtenir d'autres conseils sur la sélection du modèle ou d'appliquer la méthode moins souhaitable d'interpolation linéaire utilisant le programme ICPIN (v. le § 6.4 et l'annexe N). Dans les méthodes récentes d'essais de sol d'Environnement Canada, le programme ICPIN est l'option de repli pour l'analyse si la régression ne fonctionne pas (EC, 2004a, b et 2007).

L'homoscédasticité des résidus devrait être évaluée au moyen du *test de Levene,* décrit dans le § P.2.3 de l'annexe P, et de l'examen des graphiques des résidus en fonction des valeurs réelles et prévues. Le test de Levene indique clairement le caractère homogène ou non des données (comme dans la fig. O.2A de l'annexe O). Si les données sont hétéroscédastiques, il faudrait examiner les graphiques des résidus. Si la variance change significativement et si les graphiques des résidus présentent une dispersion nettement en fuseau ou en V (v. la fig. O.2B de l'annexe O), il faudrait répéter l'analyse à l'aide de la régression pondérée. Avant de choisir cette dernière, on devrait comparer l'erreur type de la CI p à celle de la régression non pondérée. Si les deux erreurs types diffèrent de plus de 10 %, on choisit d'abord la régression pondérée [48]. Cependant, si la différence est inférieure à 10 %, l'utilisateur devrait consulter un statisticien pour l'application d'autres modèles, sinon les données pourraient être réanalysées à l'aide de l'interpolation linéaire (moins souhaitable). On parachève cette comparaison entre la régression pondérée et non pondérée pour chacun des modèles envisagés, tout en poursuivant la sélection finale du modèle et de la régression. Certains modèles non divergents pourraient être révélateurs d'un modèle qui ne convient pas ou qui est erroné (par ex. la fig. O.2C de l'annexe O), et nous incitons de nouveau l'utilisateur à consulter un statisticien pour obtenir ses conseils sur d'autres modèles convenables.

Choix de progiciels de statistiques. — Les descriptions qui ont précédé font allusion à un progiciel de statistique polyvalent (SYSTAT), mais l'avenir pourrait nous apporter des progiciels spécialisés, conçus pour l'écotoxicologie. Par exemple, le progiciel CETIS renferme un choix extrêmement riche de modèles de régression non linéaire. Jackman et Doe (2003) ont comparé ses estimations de paramètres de toxicité à celles de SYSTAT pour de nombreux modèles. Ils ont constaté que les deux progiciels et divers modèles ont produit des estimations semblables de la CE 20 à partir d'une sélection de résultats de véritables essais de toxicité sublétale. Cependant, ils ont précisé que les résultats ont souvent varié considérablement selon différentes techniques et que des méthodes ont donné des résultats tout à fait inappropriés.

Plus précisément, Jackman et Doe (2003) signalent avoir obtenu des résultats semblables avec SYSTAT et CETIS pour 13 ensembles de données sur la toxicité sublétale à l'égard de divers organismes. Dans deux autres cas, les résultats obtenus par SYSTAT ont semblé plus convenables et, dans un autre cas, c'était l'inverse. Ces deux auteurs ont trouvé que CETIS était plus compliqué et d'un apprentissage plus difficile que les anciens progiciels de toxicologie. D'après eux, pour faire les bons choix statistiques, il faut bien comprendre les méthodes statistiques (ou disposer d'un ou de guides très détaillés). Leur recommandation est de fournir aux non-statisticiens de bons conseils sur le choix du modèle non linéaire convenable parmi tous ceux qu'offre le progiciel CETIS. Ils ont aussi fait observer que si on l'employait pour estimer des CI 50, CETIS n'estimait pas souvent des limites convenables de confiance.

48. La valeur de 10 % est uniquement fondée sur l'expérience. Des essais objectifs permettant de juger de l'amélioration due à la pondération existent, mais ils sortent du cadre du présent document. On devrait utiliser la pondération uniquement lorsqu'elle est nécessaire, la procédure pouvant introduire des complications supplémentaires à la modélisation. On devrait consulter un statisticien lorsque la pondération est nécessaire.

À mesure que davantage de progiciels de toxicologie dotés de fonctions de régression non linéaire deviendront disponibles, il importera de pouvoir compter sur les conseils d'un statisticien pour les utiliser. Il sera également souhaitable de pouvoir comparer les paramètres de toxicité estimés grâce aux nouveaux progiciels à ceux que l'on estime avec les progiciels de statistique polyvalents utilisant la méthode normalisée publiée par Environnement Canada.

6.5.9 Un nouveau programme de régression : Newtox-Logstat

Repères

- *La méthode Newtox-Logstat permet d'obtenir des estimations ponctuelles par régression. Elle a été utilisée avec succès au Canada dans les essais sur l'inhibition de la croissance chez la lentille d'eau.*
- *Le programme Newtox-Logstat offre une méthode nouvelle d'estimation ponctuelle de données quantitatives sur la toxicité sublétale, du moins sur la croissance végétale. Il offre deux modèles fondés sur la distribution de Weibull et la distribution log-normale. Il ne permet pas de modéliser les effets hormétiques, mais, à l'avenir, ses capacités pourraient être augmentées.*

Créé à l'Université technique du Danemark par les Drs K.O. Kusk et N. Nyholm, le programme d'analyse de la toxicité Newtox-Logstat s'inspire d'une méthode décrite par Andersen (1994). [Les grands principes d'une méthode similaire ont été publiés par Andersen *et al.* (1998), bien que dans un but différent. Une publication antérieure (Nyholm *et al.*, 1992) avait frayé la voie à la nouvelle méthode en décrivant les avantages de la régression non linéaire dans la résolution des difficultés statistiques posées par les données quantitatives.] Le programme a été utilisé au Canada, comme il est décrit dans le texte qui suit. Pour les chercheurs canadiens, sa source la plus commode est le Saskatchewan Research Council, en vertu d'une entente entre les Drs Kusk et Nyholm ainsi que Mary Moody[49].

La méthode Newtox-Logstat convient aux résultats quantitatifs d'essais de toxicité sublétale. Elle emploie une feuille de calcul d'Excel, et on y saisit chaque point de données et non seulement les effets moyens. Elle offre le choix entre deux modèles non linéaire fondés sur la distribution de Weibull et la distribution log-normale. Elle permet d'estimer la CI p et ses limites de confiance.

À l'origine, le programme était conçu pour les données sur les taux de croissance. Il a été utilisé avec succès au Canada par Moody (2003), pour les données sur la toxicité inhibant la croissance dans la lentille d'eau (*Lemna sp.* ; inhibition de l'augmentation du nombre de frondes et du poids sec). Moody signale que le modèle de Weibull a permis le meilleur ajustement visuel aux données. Un exemple de l'ajustement pour l'inhibition de nombre de frondes est présenté à la fig. 17.

La méthode Newtox-Logstat offre aux chercheurs canadiens une méthode de rechange pour l'estimation des paramètres de toxicité par régression, certainement pour les essais employant des algues et la lentille d'eau et probablement pour d'autres effets sur la croissance. Dans son état actuel, elle n'est pas capable de prendre en charge le phénomène de l'hormèse dans le modèle. Ses concepteurs ont proposé de régler arbitrairement les effets hormétiques à 0 % d'inhibition, pour les besoins de la modélisation. Moody (2003) a constaté que, en général, l'hormèse ne gênait pas l'analyse, mais elle a écarté les données « hormétiques » quand elles faisaient problème[50].

49. Mme Mary Moody, chercheure, Environment and Mineral Branch (Direction générale de l'environnement et des minéraux), Saskatchewan Research Council, 125 - 15, Innovation boulevard, Saskatoon S7N 2X8 (moody@src.sk.ca).

50. À partir d'essais d'inhibition de la formation des frondes, Moody (2003) a comparé 23 paramètres de toxicité estimés par régression à des paramètres homologues obtenus après lissage et interpolation (par le programme ICPIN). Les quotients entre les paramètres (estimés par interpolation / estimés par régression) ont révélé une similitude complète (quotient moyen de 102 % ; quotient médian de 96 %). Cependant la diversité était grande dans les comparaisons individuelles. Les quotients variaient de 42 à 195 %, avec un écart type de 39 %. En supposant que les paramètres de toxicité estimés par régression sont plus réalistes, ils représentent une amélioration méthodologique appréciable. La comparaison des poids secs des plantes a présenté une similitude plus grande entre les paramètres de toxicité estimés par les deux méthodes (écart type de 20 %), mais, dans 7 cas, l'interpolation n'a pas permis d'estimer le paramètre de toxicité ou n'a pas permis de déterminer les limites de confiance.

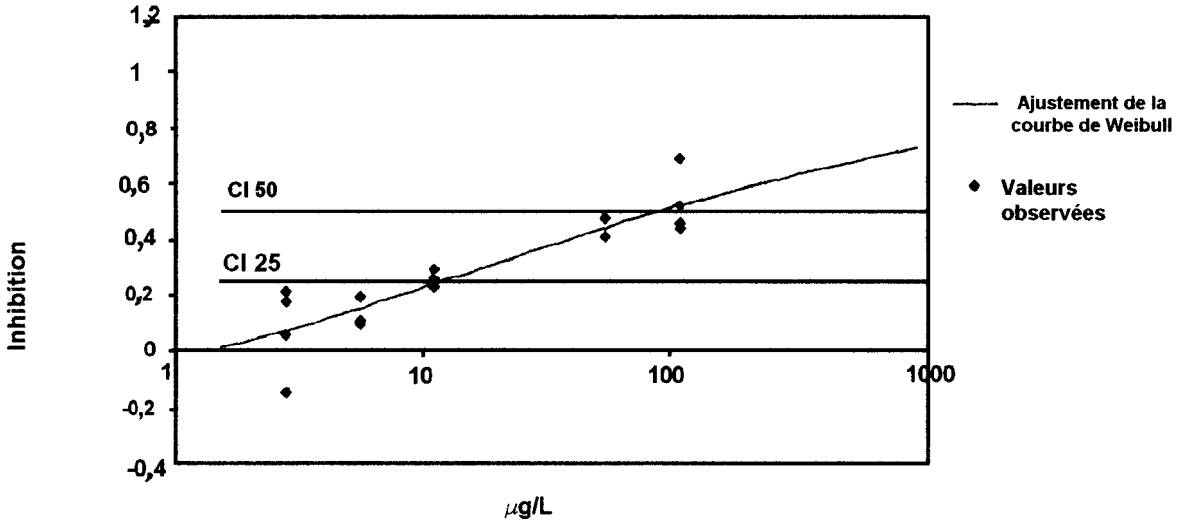


Figure 17. — Effet du cadmium sur l'inhibition de la croissance des frondes chez *Lemna minor* (d'après Moody, 2003). La courbe ajustée se fonde sur un modèle de Weibull utilisant les méthodes de Kusk et de Nyholm inspirées par Anderson (1998) du progiciel Newtox-Logstat.

6.5.10 Modèles linéaires généraux

Repères

- *Les modèles linéaires généraux (GLM) et les modèles linéaires généralisés (GLIM) sont des catégories générales de modèles statistiques comprenant de nombreuses techniques statistiques bien connues.*

- *La catégorie la plus nombreuse est celle des GLIM, qui englobe diverses distributions, notamment la normale, l'exponentielle, la logistique, celle de Poisson et la distribution probit. L'approche pourrait s'appliquer aux études des variables quantiques ou des variables continues telles que le poids.*

- *Actuellement, ces notions restent du domaine de la statistique, mais elles ont été affinées pour servir dans celui de la toxicologie. Des développements avantageux pourraient être : l'obtention d'un progiciel unique, permettant d'analyser diverses catégories de résultats ; le transfert de connaissances et de techniques entre les modèles ; l'utilisation de meilleures méthodes mathématiques plutôt que de techniques inexactes ; la comparaison de l'ajustement de divers modèles. Cependant, les non-statisticiens trouveraient probablement les progiciels existants difficiles à utiliser.*

L'expression *modèles linéaires généraux (GLM)* n'englobe pas une technique particulière, mais, plutôt, une catégorie d'approches ou de modèles. Les modèles possèdent une variable dépendante unique, qui est fonction d'une ou de variables indépendantes. Ainsi, la régression linéaire simple entre dans cette catégorie, mais on ne devrait pas se représenter les GLM comme étant limités aux régressions. Entrent également dans les GLM les modèles tels que l'analyse de variance et l'analyse de covariance, qui ne pourraient pas être considérées comme des modèles « linéaires ». Les statisticiens feraient remarquer que ces méthodes sont « linéaires » parce que leurs paramètres entrent dans le modèle d'une façon linéaire. Gad (1999) donne un exemple dans lequel la méthode de GLM du logiciel SAS est mise à contribution pour effectuer une analyse de variance classique de données toxicologiques typiques (poids des reins en fonction de plusieurs doses).

Ainsi, les expérimentateurs devraient s'attendre à trouver beaucoup de techniques d'analyse particulière dans la catégorie générale des GLM. Les GLM ne constituent pas un logiciel s'appliquant simplement et bêtement à un ensemble de données. La plupart des biologistes ou des toxicologues auraient besoin de la participation directe d'un statisticien pour appliquer ces techniques à leur travail. Des GLM ont été décrits par Searle (1971).

6.5.11 Modèles linéaires généralisés

L'expression *modèles linéaires généralisés (GLIM)* représente une catégorie plus large encore de modèles mathématiques qui englobe les GLM, dont nous venons de parler. La catégorie a parfois été appelée « modèles interactifs linéaires généralisés » (*generalized linear interactive models*), d'où l'abréviation GLIM.

Leurs grandes capacités ont d'abord fait des GLIM d'utiles moyens didactiques. Mais la recherche et l'avènement d'ordinateurs puissants les ont propulsés sur le devant de la scène des développements et de l'actualité statistiques. Tous les GLIM ont en commun la même approche mathématique, mais la catégorie pourrait englober diverses techniques particulières. Les techniques elles-mêmes pourraient intéresser plus directement la toxicologie appliquée que les notions mathématiques abstraites sur lesquelles les GLIM reposent. Dobson (2002) a rédigé une introduction à la question, tandis que McCullagh et Nelder (1994) ont conçu un manuel plus détaillé pour les statisticiens et « les biologistes à l'aise avec les chiffres ».

Diverses distributions mathématiques bien connues se classent sous la rubrique générale des GLIM, y compris les distributions normale, exponentielle, logistique, la distribution probit et celle de Poisson. On peut les décrire mathématiquement de façon à ce qu'un effet dans n'importe quelle d'entre elles puisse être lié, par une fonction, à une ou à plusieurs variables indépendantes. L'effet pourrait être quantique (dénombrements, mortalités, proportions) ou être une variable continue telle que le poids. Il existe une méthode commune de calcul des estimations des paramètres. Un chercheur pourrait utiliser les GLIM pour évaluer la dépendance d'un effet à l'égard d'une seule variable indépendante telle que la concentration (par régression) ou une structure plus complexe de variables indépendantes tel qu'un traitement collectif (analyse de variance) ou des traitements et des covariables (analyse de covariance).

Dans une série de communications, Bailer et Oris (1993 ; 1994 ; 1997) et leurs associés (Bailer *et al.*, 2000a, b) ont donné un appui enthousiaste à l'emploi des GLIM dans les analyses toxicologiques. Ils ont montré que leur modèle général de régression peut s'ajuster à différents effets, qu'ils soient dénombrés, dichotomiques ou continus. La régression peut servir à estimer la CI p et à contourner les problèmes conceptuels existant dans le programme informatique ICPIN (Bailer et Oris, 1997), étant « supérieure [au programme ICPIN] pour ce qui concerne le biais, l'erreur quadratique moyenne et le taux de couverture » (Bailer *et al.*, 2000b). Bailer et Oris (1994) font observer que le logiciel permettant d'ajuster les GLIM est d'accès facile (par ex. la macro GLIM de la procédure NLIN du logiciel SAS. Dans les publications antérieures, les limites de confiance ne sont pas estimées, mais Bailer et Oris (1997) énumèrent des options que l'on pourrait élaborer en les dotant d'une base mathématique défendable.

L'emploi des GLIM et des GLM qui en font partie confère plusieurs avantages, comme suit :

- Un seul progiciel peut remplacer l'éventail de programmes dont on a besoin pour analyser des effets non normaux et linéaires.

- L'expérimentateur peut se servir des mêmes connaissances générales des types de modèles faisant partie des GLIM (par ex. signification, qualité de l'ajustement, test d'hypothèses).

- On peut cesser d'utiliser les approches qui englobent des raccourcis inexacts et des techniques à des fins particulières datant de l'époque antérieure à l'avènement de l'ordinateur et adopter plutôt de meilleures méthodes mathématiques.

- La comparaison de l'ajustement aux données sur la relation dose-effet de diverses distributions (par ex. probits, logistique, Gompertz) est simple.

En même temps, les GLIM souffrent de limites et d'inconvénients. Bien qu'il existe un progiciel

autonome pour les GLIM, les biologistes pourraient en trouver l'utilisation difficile. L'achat d'un progiciel de plus d'envergure tel que SAS permettrait d'employer les GLIM, mais les utilisateurs devraient apprendre à faire appel aux techniques appropriées et à les maîtriser. Manifestement, les GLIM sont utiles à la recherche en toxicologie, mais les essais de toxicité en routine ou à des fins réglementaires suivront probablement des pistes toutes faites, telles les conseils donnés en matière de statistique dans les méthodes d'Environnement Canada.

6.5.12 Reparamétrage

Cette approche à l'analyse des données sur la toxicité découle du désir d'estimer les paramètres de toxicité et leurs limites de confiance et de les exprimer en fonction d'une concentration particulière (CE 50, CI 25), bien que les essais de toxicité aient été mis au point avec la concentration comme *variable indépendante.* Le degré d'effet était, en réalité, la variable dépendante. Pourtant, on se sert d'un degré fixe d'effet pour calculer le paramètre de toxicité en unités de concentration. Cette « inversion » de la régression pour choisir un paramètre de toxicité entraîne des complications statistiques, décrites dans le § 9.4. Une façon de contourner le problème est le reparamétrage pour créer un modèle renfermant le paramètre de toxicité auquel on s'intéresse. L'approche a été adoptée par Stephenson *et al.* (2000), puis modifiée dans les méthodes récentes d'Environnement Canada (§ 6.5.8).

Repères

- *Dans les essais d'écotoxicité, l'effet mesuré est la variable dépendante. Pour calculer le paramètre de toxicité, cependant, on se sert d'un degré fixe d'effet, comme si c'était la variable indépendante, afin de calculer la concentration correspondante de matière toxique (le paramètre de toxicité). Cela entraîne une « inversion » de la relation.*
- *Le reparamétrage implique une modification du modèle statistique décrivant la relation, de façon à intégrer le paramètre de toxicité (la CI p, disons) et ses limites de confiance en tant que variables à estimer par le modèle. C'est ce que font les techniques de régression linéaire et non linéaire d'Environnement Canada (annexe O).*
- *Cette façon de faire pourrait diminuer les performances des modèles, avec, comme pénalité, la nécessité d'accroître le nombre de répétitions pour obtenir des résultats satisfaisants.*
- *D'autres auteurs ont publié des approches à la modélisation non linéaire, et nous en exposons brièvement des exemples.*

Le reparamétrage commence par un modèle statistique usuel, tel que le modèle de régression non linéaire. Si on doit estimer la CI 25, on « reparamètre » l'équation de régression en y incluant la CI 25 comme paramètre. Cela permet d'estimer directement la CI 25 et ses limites de confiance sans devoir employer les techniques de régression inverse.

Cette opération comporte des inconvénients. L'analyse statistique, notamment, pourrait fonctionner moins bien. Par exemple, le modèle hormético-logistique (§ 6.5.8) s'est révélé très sensible au choix de l'algorithme d'optimisation. Il est donc souhaitable d'éprouver un nombre de concentrations plus grand que la normale.

Les premiers, Bruce et Versteeg (1992) ont décrit le reparamétrage dans un excellent exposé de l'emploi de la régression non linéaire sur les données quantitatives sur la toxicité. Ils ont éprouvé la méthode sur des essais de toxicité sublétale avec l'algue *Pseudokirchneriella subcapitata,* des têtes-de-boule et des crustacés du sous-ordre des mysidés. Les courbes résultantes des effets mesurés à différents logarithmes de la concentration semblaient des ajustements lissés. Le programme a alors reparamétré l'équation de la droite ajustée pour estimer la CI p logarithmique et ses limites de confiance, pour n'importe toute valeur choisie de *p*. Ayant fondé leur modèle sur « sur une courbe en forme de S dérivée de la fonction de répartition d'une loi normale », Bruce et Versteeg (1992) divulguent le code pour l'exécution de l'analyse avec le logiciel SAS. Un autre exemple est donné par Andersen *et al.* (1998).

Cette marche à suivre a été intégrée dans les modèles offerts pour les nouvelles méthodes d'essai d'Environnement Canada (EC, 2004a, b et 2007), décrites dans le § 6.5.8 et l'annexe O.

6.5.13 Autres exemples de tentatives de régression

Nous exposerons les méthodes de régression utilisées par d'autres auteurs. Elles semblent prometteuses, mais leur application exige que l'on s'y connaisse.

Slob (2002) a décrit une famille de modèles non linéaires, semblables à ceux dont nous venons de discuter. Les analyses sont effectuées par un progiciel d'utilisation facile, appelé PROAST, disponible dans l'institut néerlandais où travaille Slob. L'une des qualités des régressions est de permettre la détermination de la *dose critique* (*Critical Effect Dose* [*CED*]), qui est reliée à un degré négligeable ou acceptable d'effet sur les organismes en expérience.

Andersen *et al.* (1998) ont recommandé une régression non linéaire généralisée pour estimer la CI p et ses limites de confiance. Pour choisir une fonction particulière de régression, ils ont utilisé un graphique sur lequel les données étaient portées. Pour leur analyse, ils ont fusionné des routines numériques usuelles, y compris un codage en FORTRAN 90. Leur méthode a fait appel à la « variance et à la covariance non homogènes empiriques pour l'estimation de la courbe dose-réponse ». Une version tournait sur la plate-forme Windows 95.

Scholze *et al.* (2001) ont utilisé 10 fonctions sigmoïdes différentes de régression, les mieux connues étant les régressions probit, logit, de Weibull, logit généralisées et trois options des fonctions Box-Cox. Toutes les fonctions ont été ajustées à un ensemble particulier de données, et on a retenu le meilleur ajustement au moyen de deux étapes de tests (des résidus, puis de la qualité de l'ajustement). On a estimé les limites de confiance par la méthode bootstrap. La méthode a été confirmée grâce à une prédiction remarquable de la toxicité totale d'un mélange de 14 substances à différents modes d'action. L'effet prévu sur l'inhibition de la luminescence bactérienne était 36 %, ce qui est presque identique aux 39 % effectivement observés.

Moore et Caux (1997) ont appliqué cinq modèles « génériques » à des données quantiques et quantitatives. Les meilleurs ajustements ont habituellement été obtenus par une équation logistique à trois paramètres, dont un avait une pente fortement inclinée. Ils ont aussi essayé trois modèles logistiques, un modèle probit à deux paramètres et un modèle de Weibull à deux paramètres. Ils ont exclu les polynômes d'ordre supérieurs, peu plausibles du point de vue biologique. Leur progiciel, employant une feuille de calcul, se servait du logarithme de la concentration et donnait un maximum de vraisemblance convenable avec chaque modèle (Caux et Moore, 1997). Le logiciel a livré une estimation de qualité de l'ajustement, des graphiques établis avec les données observées et dotés d'une courbe ajustée ainsi que les CE p ou les CI p correspondant aux valeurs petites et grandes de *p*. Parmi 198 ensembles de données sur la toxicité sublétale, ils en ont choisi 65, dans lesquels la relation dose-effet était convenablement monotone et un effet au moins était partiel. Ils ont analysé les 65 ensembles à l'aide de leur méthode et ils ont allégué un ajustement adéquat dans une quarantaine de cas

Baird *et al.* (1995) ont prétendu qu'il suffisait de seulement deux modèles paramétriques non linéaires pour expliquer divers résultats sur la toxicité. À l'aide d'un *modèle logistique dose-réponse* et d'un *modèle puissance*, ils ont ajusté de 77 à 100 % des essais quantitatifs de toxicité sublétale effectués avec des ménés, des oursins, des ormeaux et des laminaires géantes (ou macrocystes). Le modèle puissance avait la forme $y = bx^c$ et il s'ajustait à des droites ainsi qu'à des distributions concaves ou convexes vers le haut. Cependant, leur validation a été incertaine, du fait que leurs données hypothétiques possédaient des étendues arithmétiques fortement irréalistes et que les données logarithmiques ont été analysées à l'aide de concentrations arithmétiques. Les données sur la laminaire géante ont été ajustées et représentées graphiquement sous forme de courbe, mais, apparemment, elles auraient épousé la forme d'une droite si on avait utilisé une bonne échelle logarithmique de concentration.

6.6 Seuils estimés par régression

Repères

- *Un mouvement international aspire à la mise au point des méthodes qui permettraient d'estimer la concentration vraie ou absolue sans effet pour une population d'organismes. Ce serait une valeur théorique, qui devrait être estimée par des techniques de régression.*

- *Aux Pays-Bas, on a construit de tels modèles pour estimer la concentration sans effet ou la dose critique.*
- *Les modèles « en bâton de hockey » permettent d'estimer un tel seuil d'effet. Le long manche est constitué de la régression normale de la relation concentration-effet, tandis que la lame représente l'arrière-plan d'effets normaux. L'intersection des deux droites est censée représenter un seuil.*

En Europe et ailleurs, un mouvement aspire à la construction de modèles de la toxicité permettant d'estimer la « vraie » *concentration sans effet* (OECD, 2004). Cette estimation serait une concentration absolue sans effet, qui est un paramètre de la population et non de l'échantillon soumis à l'essai (Anonyme, 1994). Le but serait approché par des techniques de régression et non par un test d'hypothèse qui estime la concentration sans effet observée (CSEO) dans un échantillon, plutôt que la vraie concentration. Nous donnons des exemples de ces avancées européennes dans le § 6.6.2.

6.6.1 Seuils estimés par le modèle en bâton de hockey

On peut modéliser le seuil de toxicité sublétale quantitative au moyen d'un modèle dit « en bâton de hockey ». C'est un modèle linéaire de régression puisque deux droites sont ajustées aux résultats de l'essai. La droite évoquant le manche s'applique à la relation dose-effet habituelle, tandis que ce qui tient lieu de lame serait une droite parallèle à l'axe des concentrations (fig. 18). Zajdlik (1996) a franchi les étapes mathématiques de l'ajustement d'un tel modèle et, d'après lui, cette opération n'est pas difficile.

La description de Zajdlik confère beaucoup d'attrait à l'approche, qui permet d'estimer un seuil apparemment objectif et significatif d'effet se situant à l'intersection des deux droites. Il fait remarquer des inconvénients potentiels tels que le problème général de l'ajustement d'une relation dose-effet incurvée. Parfois, un toxique particulier pourrait ne pas manifester un seuil d'effet (les faibles concentrations agiraient simplement plus tard). Zajdlik (1996) mentionne qu'il pourrait être plus coûteux d'effectuer l'expérience pour ce type d'analyse, mais que les coûts seraient équilibrés par les avantages d'une estimation objective d'une concentration sans danger pour l'environnement.

La méthode discutée par Zajdlik (1996) a été utilisée par d'autres auteurs. Un excellent exemple concerne l'incidence de lésions hépatiques chez les poissons benthiques, en fonction de la présence d'aromatiques polycycliques (HAP) dans les sédiments (Horness *et al.*, 1998). La représentation graphique (fig. 18) montre l'incidence naturelle des lésions distribuées horizontalement le long de l'un des segments du « bâton de hockey », sur un intervalle de faibles concentrations logarithmiques d'HAP. Puis, on constate un changement brusque, le second segment de la régression montrant une augmentation linéaire des lésions coïncidant avec les concentrations logarithmiques supérieures. Les ajustements semblent convenables, bien que les intervalles de confiance soient plutôt larges dans les graphiques A et D. L'intervalle de confiance ne figure pas dans le graphique B, mais il pourrait être large. Cependant, l'estimation d'un seuil apparent d'effet toxique à l'intersection des deux segments semble un élément d'information très utile.

Pour cette analyse, Horness *et al.* (1998) ont traité les deux segments (droites) comme une seule fonction discontinue, définie par une régression simple. Les concentrations d'HAP ont été transformées en logarithmes avant l'analyse, bien que cela ait pu avoir été effectué dans le cadre des calculs. Horness *et al.* font remarquer que des techniques numériques itératives d'estimation des paramètres de régression non linéaire sont de plus en plus souvent offerts dans les progiciels commerciaux usuels, et ils ont utilisé le progiciel de statistique JMP® de SAS.

L'utilité potentielle des « paramètres de toxicité estimés grâce au modèle de bâton de hockey » est éloquemment prouvée par Beyers *et al.* (1994), qui ont estimé des seuils de toxicité 2 à 4 fois plus bas que la CSEO estimée par un test d'hypothèse. Ils ont étudié la toxicité de pesticides pour le poisson, et leurs ajustements du modèle en bâton de hockey semblent satisfaisants. Ils ont aussi utilisé le logiciel de statistique développé par SAS.

6.6.2 Estimation de la concentration sans effet par régression

La régression non linéaire devrait être utilisable pour estimer des seuils d'effet toxique, approche qui a été

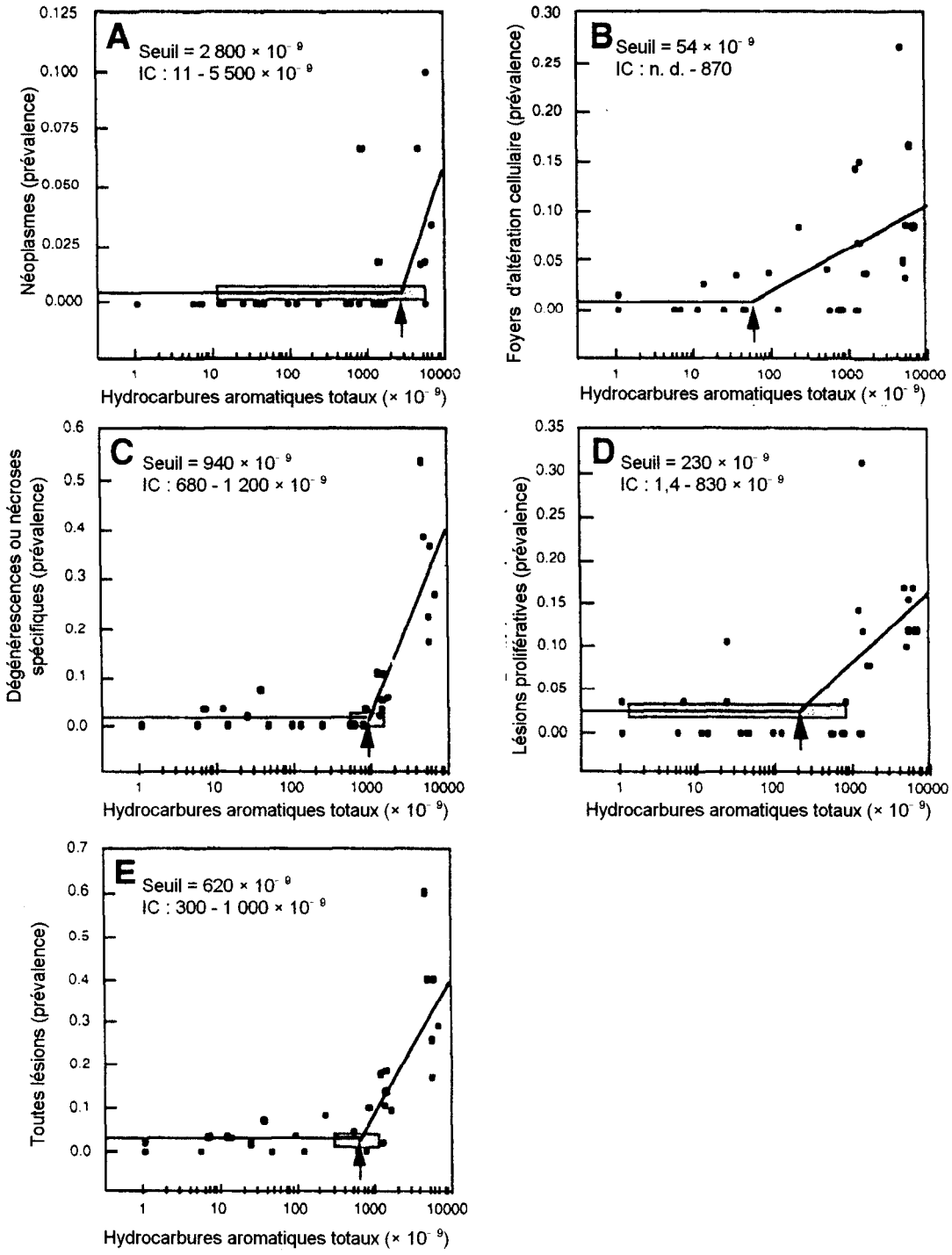

A
Seuil = 2 800 × 10⁻⁹
IC : 11 - 5 500 × 10⁻⁹
Néoplasmes (prévalence)
Hydrocarbures aromatiques totaux (× 10⁻⁹)
B
Seuil = 54 × 10⁻⁹
IC : n. d. - 870
Foyers d'altération cellulaire (prévalence)
Hydrocarbures aromatiques totaux (× 10⁻⁹)
C
Seuil = 940 × 10⁻⁹
IC : 680 - 1 200 × 10⁻⁹
Dégénérescences ou nécroses spécifiques (prévalence)
Hydrocarbures aromatiques totaux (× 10⁻⁹)
D
Seuil = 230 × 10⁻⁹
IC : 1,4 - 830 × 10⁻⁹
Lésions prolifératives (prévalence)
Hydrocarbures aromatiques totaux (× 10⁻⁹)
E
Seuil = 620 × 10⁻⁹
IC : 300 - 1 000 × 10⁻⁹
Toutes lésions (prévalence)
Hydrocarbures aromatiques totaux (× 10⁻⁹)

Figure 18. — Exemples de régression en bâton de hockey (d'après Horness *et al.,* 1998). Les graphiques représentent les données pour certains types de lésions hépatiques chez des soles anglaises capturées dans des localités de la côte du Pacifique. Les échelles verticales représentent la prévalence chez le poisson. Les axes horizontaux mesurent la teneur en hydrocarbures aromatiques totaux dans les sédiments séchés du fond, prélevés dans les mêmes localités. Les concentrations seuils sont indiquées par des flèches, tandis que les bandes ombrées représentent les intervalles de confiance (IC).

tentée aux Pays-Bas. Slob (2002) a montré l'emploi d'une famille de régressions non linéaires pour déterminer la *dose critique,* reliée à un effet négligeable chez les organismes en expérience (v. le § 6.5.13).

De même, Kooijman et Bedaux (1996) offrent une description et un logiciel pour estimer le paramètre de toxicité sublétale appelé « concentration sans effet (CSE). Leur programme est conçu principalement pour l'analyse des méthodes d'essais de toxicité sublétale déterminer la *dose critique,* reliée à un effet négligeable chez les organismes en expérience (v. le § 6.5.13).

De même, Kooijman et Bedaux (1996) offrent une description et un logiciel pour estimer le paramètre de toxicité sublétale appelé « concentration sans effet (CSE). Leur programme est conçu principalement pour l'analyse des méthodes d'essais de toxicité sublétale publiées par l'OCDE sur la croissance des poissons, la reproduction de la daphnie ou puce d'eau (*Daphnia*) et la croissance des algues. Ils mentionnent que le programme peut aussi produire des analyses des données quantiques sur la mortalité (CL 50), des concentrations efficaces (CE 50) et des temps efficaces (TE 50), tous accompagnés de leurs limites de confiance. Ces allégations n'ont pas été validées pour le présent document d'Environnement Canada.

Kooijman et Bedaux (1996) ont glissé la disquette du programme informatique (programme DEBtox, pour *Dynamic Energy Budget,* ou bilan énergétique dynamique) dans une monographie. Publié en 1996, le programme tournait sur la plate-forme Windows 3.1 ou 95. Des versions plus récentes fonctionnent sous Windows et Unix et sont offerts dans Internet à l'adresse suivante : www.bio.vu.nl/thb/deb/deblab/. Les fonctions du programme ont été décrites en détail, récemment, dans un guide de l'OCDE (OECD, 2004). Le programme semble bien conçu, clair et facile à utiliser. Les données fournies à titre d'exemple sont traitées facilement par le programme ; ce dernier estime les paramètres de toxicité et l'information à l'appui, mais il ne donne pas d'indications des modèles et des méthodes utilisées pour obtenir les réponses [51]. Le programme offre des graphiques imprimables si on le désire. Malheureusement, les concentrations ont été représentées sur une échelle arithmétique qui donne au lecteur une impression déformée des asymptotes, des seuils apparents et de l'allure générale des courbes.

L'approche fondée sur la concentration sans effet est également intégrée dans une fonction mathématique de l'inhibition de la croissance de la population dans les essais employant des algues (Kooijman *et al.*, 1996). On dit que l'équation fonctionne bien, parvenant à l'efficacité des analyses logistiques, log-normales ou par la méthode de Weibull (N. Nyholm, Université technique du Danemark à Lyngby, communication personnelle, 2001).

Les avantages de l'approche fondée sur la concentration sans effet sont évidents. Elle utilise des méthodes statistiques appropriées, c'est-à-dire l'ajustement d'une régression. Elle satisfait la demande d'un paramètre de toxicité représentant le seuil d'effet, censément l'absence d'effet.

51. Une tentative de saisie de données nouvelles a échoué. L'opérateur canadien a réussi à saisir des chiffres dans certaines positions du tableau de données initiales, mais il n'a pas réussi à découvrir quelles parties du tableau devaient recevoir les données sur les concentrations, le temps, le nombre d'organismes en expérience et l'effet. Il n'a pas pu obtenir de conseils sur ces points.

Tests d'hypothèse(s) pour déterminer la concentration sans effet observé (CSEO) et la concentration avec effet minimal observé (CEMO)

7.1 Pertinence générale pour les essais d'écotoxicité

Repères

- *Le test d'hypothèse détermine les différences statistiquement significatives entre les résultats obtenus avec le témoin et les résultats obtenus à chaque concentration.*
- *C'est une approche qui convient aux essais à concentration unique comme aux essais de surveillance.*
- *Dans un essai à plusieurs concentrations, le test d'hypothèse(s) permet d'identifier la CSEO et la CEMO.*
- *L'estimation de la CSEO et de la CEMO est une option dans certaines méthodes d'essai de la toxicité sublétale publiées par Environnement Canada. Cependant, elle ne représente pas un bon paramètre toxicologique dans les essais à plusieurs concentrations, pour plusieurs raisons, les suivantes notamment :*
  - *Les paramètres de toxicité sont définis statistiquement plutôt que biologiquement ; une plus grande variabilité dans l'essai conduit à des valeurs plus élevées de la CSEO et de la CEMO.*
  - *En dépit de l'impression qu'elle donne, la CSEO ne représente pas nécessairement une concentration sans danger (inoffensive) dans l'environnement.*
  - *Les paramètres de toxicité peuvent seulement être des concentrations qui ont effectivement été éprouvées et, de ce fait, ils sont exposés aux effets du hasard ou des plans d'expérience.*
  - *Les calculs donnent une paire de concentrations plutôt qu'un seul paramètre de toxicité.*
  - *On ne peut calculer aucune limite de confiance.*
- *La moyenne géométrique de la CSEO et de la CEMO peut servir à représenter un paramètre unique de toxicité que l'on devrait appeler la* concentration avec effet de seuil observé *(CESO). Elle présente les mêmes inconvénients que la CSEO et la CEMO.*

7.1.1 Essais à concentration unique

Le test d'hypothèse(s) est le mode opératoire normalisé pour les essais de toxicité dans lesquels on a employé des répétitions d'une concentration et un témoin (par ex. des échantillons de sédiment d'un emplacement). C'est une méthode statistique appropriée, et il n'y en a pas d'autres. Les techniques disponibles sont décrites dans la section 3.

On peut comparer un traitement et un témoin à l'aide du *test t*. Il ne faut pas répéter les tests *t* multiples sur un ensemble d'échantillons au lieu d'effectuer un test de comparaisons multiples. Il existe des modifications particulières du test *t* (annexe P.4.4).

D'ici la fin de la section, il sera question d'essais comportant au moins deux concentrations ou deux collections d'échantillons.

7.1.2 Essais à plusieurs concentrations

Le principal paramètre de toxicité que nous recommandons d'estimer est une estimation ponctuelle telle que la CI 25. Le test d'hypothèse(s) est considéré comme secondaire. Cependant, plusieurs méthodes d'Environnement Canada permettent l'emploi du test d'hypothèse(s) si on le désire. En conséquence, nous exposons ci-dessous des méthodes en ce sens, puisqu'elles pourraient s'appliquer à une situation particulière et, aussi, permettre l'évaluation de travaux antérieurs ayant utilisé cette méthode statistique.

Les variables estimées dans le test d'hypothèse(s) seraient la *concentration sans effet observé (CSEO)* et la *concentration avec effet minimal observé (CEMO)*. La méthode habituelle de détermination de la CSEO et de la CEMO consiste à comparer, statistiquement, l'effet observé chez les témoins aux effets observés à chaque concentration expérimentale (v. les § 7.4 et 7.5). Le test d'hypothèse(s) est souvent utilisé, en partie

parce qu'il s'agit de méthodes bien établies. L'analyse de variance et les méthodes non paramétriques sont très accessibles, relativement faciles à utiliser et robustes à l'égard des données irrégulières. Cependant, de plus en plus de publications signalent les nombreuses carences de l'approche fondée sur le test d'hypothèse(s) (Suter *et al.*, 1987 ; Miller *et al.*, 1993 ; Pack, 1993 ; Noppert *et al.*, 1994 ; Chapman, 1996 ; Chapman *et al.*, 1996b ; Pack, 1996 ; Suter, 1996 ; Moore et Caux, 1997 ; Bailer et Oris, 1999 ; Andersen *et al.*, 2000 ; Crane et Newman, 2000 ; Crane et Godolphin, 2000). Parmi les douteurs, on trouve des écotoxicologues et des statisticiens canadiens (Miller *et al.*, 1993). Voici quelques-unes des limites de cette méthode :

- La CSEO et la CEMO ne peuvent avoir que des valeurs égales aux concentrations effectivement utilisées. Comme ces concentrations ont été décidées par l'expérimentateur, les paramètres de toxicité pourraient être exposés à des influences fortuites, au caprice ou à la manipulation.

- La CSEO et la CEMO sont particulièrement sensibles à la variabilité interne de l'essai, puisqu'elles dépendent de la détermination d'une différence statistiquement significative par rapport à l'effet observé chez le témoin. Un essai soigneusement mené et ayant abouti à un résultat précis abaisserait la CSEO, tandis qu'un essai aux résultats très variables entraîneraient une CSEO plus élevée. Ainsi, les paramètres de toxicité CSEO et CEMO ne correspondent à aucun point particulier de la courbe dose-effet.

- Les valeurs de la CSEO et de la CEMO dépendent dans une certaine mesure de la méthode d'analyse statistique employée.

- La puissance statistique de l'analyse de variance et du test de comparaisons multiples est souvent faible, en raison du nombre relativement peu élevé de répétitions. Moins de répétitions entraînent une CSEO plus forte, de sorte qu'il pourrait y avoir une incitation soit à réduire, soit à augmenter le nombre de répétitions selon l'orientation que l'on entend donner au programme.

  [On pourrait remédier à cette situation, en s'attachant davantage aux caractéristiques de la puissance dans le plan d'expérience (§ 7.23). On pourrait exiger qu'il soit démontré qu'un essai possédait une puissance suffisante pour, disons, déceler un effet de 25 %.]

- Ne pouvant calculer aucune limite de confiance pour le paramètre de toxicité, on ne peut pas comparer statistiquement différentes CSEO.

- La CSEO, en raison de l'attrait de cette appellation pour le commun des mortels, pourrait être confondue avec une concentration « inoffensive » par les non-toxicologues, même si cette concentration peut correspondre à des effets appréciables.

- En estimant une CSEO, on tend à s'opposer quelque peu à une règle fondamentale de la méthode scientifique, parce qu'on tente de « prouver » une hypothèse nulle de l'absence d'effet.

En toute justice, on devrait faire remarquer que ces limitations ne sont pas propres au test d'hypothèse(s). La plupart se retrouvent dans d'autres méthodes d'analyse des résultats des essais. Par exemple, les intervalles classiques de confiance de la CI 25 et de la CE 25 dépendent de la justesse de l'hypothèse utilisée par le modèle pour les produire. Il est même rare que l'on reconnaisse cette hypothèse, sans parler de la tester.

L'importance de la précision des résultats et du choix de la méthode statistique est révélée par un exemple donné par Crane et Godolphin (2000). Ceux-ci présentent des données hypothétiques concernant un essai de toxicité létale sur le même effluent par le laboratoire A, qui a obtenu des résultats précis, et le laboratoire B, qui a obtenu des résultats variables. L'analyse par le test de Dunnett (actuellement le test de comparaisons multiples le plus employé) donne une CSEO de 2,2 % pour laboratoire A et de 22 % pour laboratoire B. Le choix d'autres tests statistiques donne une CSEO estimée très variable, qui va de 1,0 à 10 %, dans le cas des résultats du laboratoire A, et de 2,2 à 46 %, dans le cas du laboratoire B (v. le § 7.5.1).

On peut donner d'autres exemples particuliers. Suter *et al.* (1987) ont montré que l'estimation de la CSEO et

de la CEMO ne donnait de paramètres de toxicité satisfaisants. Lorsque des études de la toxicité sublétale pour le poisson ont été analysées par régression non linéaire, une comparaison avec les résultats du test d'hypothèse(s) a montré que les moyennes géométriques de la CSEO et de la CEMO (les *concentrations avec effet de seuil observé [CESO]* : v. le texte qui suit) étaient associées à des effets de 12 % sur l'éclosion, de 19 % sur la survie des larves et de 20 % sur la survie des parents et le poids des larves, de 35 % sur la masse d'un œuf et de 42 % sur la fécondité. Ce sont des effets relativement puissants, qui montrent certainement que les CESO peuvent être éloignées d'un véritable seuil d'effet. Une analyse semblable des résultats de 14 essais de toxicité sublétale a montré que la CSEO (et non la CESO) correspondait à des effets sublétaux qui variaient de 3 à 38 % (moyenne géométrique de 14 %) [Crane et Newman, 2000].

Pour ce qui concerne l'attrait de l'appellation *concentration sans effet observé* (CSEO), on relève, dans le compte rendu de la réunion de Québec, que le test d'hypothèse(s) est une désignation attrayante du point de vue réglementaire et gestionnel, parce qu'il donne l'impression de répondre à la question que l'on se pose sur la toxicité ou non d'une concentration donnée dans l'environnement (Miller *et al.*, 1993). Les écotoxicologues sont conscients que toute réponse à cette question par la détermination de la CSEO et de la CEMO pourrait être erronée, en raison des problèmes que nous venons d'exposer.

7.1.3 Expression des résultats sous forme de seuil

On calcule souvent la moyenne géométrique de la CSEO et de la CEMO afin de n'avoir qu'un seul nombre à manipuler plutôt que deux. Une désignation recommandée pour cette moyenne géométrique est celle de *concentration avec effet de seuil observé (CESO)*. Ici, *seuil*, conformément au dictionnaire, signifie point à partir duquel un effet commence à se manifester. Une telle valeur peut être utilisée et signalée, en reconnaissance du fait qu'elle représente une estimation arbitraire d'un seuil d'effet qui pourrait se situer n'importe où dans l'intervalle entre la CEMO et la CSEO et qui serait assujetti à toutes les incertitudes de ces valeurs (§ 7.1.2).

On s'est servi de l'expression *concentration maximale acceptable de toxique (CMAT),* principalement aux États-Unis (sous l'abréviation *MATC*), comme d'un paramètre empirique de la toxicité dans les essais de toxicité sublétale couvrant le cycle vital des sujets d'expérience. L'expression a été galvaudée par divers auteurs, qui lui ont prêté les significations suivantes : *a*) moyenne géométrique de la CSEO et de la CEMO ; *b*) CSEO ; *c*) valeur indéterminable, située entre la CSEO et la CEMO ; *d*) intervalle allant de la CSEO à la CEMO. Dans les publications récentes on tend à abandonner cette CMAT ou MATC galvaudée en lui préférant les expressions *CSEO* et *CEMO* ; nous recommandons l'emploi de l'abréviation *CESO* ou *concentration avec effet de seuil observé*. Les estimations ponctuelles conviennent davantage à la détermination de seuils (section 4).

7.2 Particularités du plan d'expérience dans le test d'hypothèse(s)

Repères

- *En analyse de variance, l'emploi de mesures non répétées, comme si c'étaient des répétitions, pourrait entraîner des erreurs majeures.*
- « Erreur α » *ou* « erreur de première espèce » *signifient* « faux positif »*, c'est-à-dire le fait de conclure à l'existence d'une différence entre des traitements alors que, en réalité, il n'existe pas de différence réelle.* « Erreur β » *ou* « erreur de seconde espèce » *signifient l'acceptation d'une hypothèse nulle de l'absence de différence bien qu'une différence réelle existe.*
- *La plupart des expérimentateurs fixent le niveau de signification* (**α**) *à la probabilité* (p) *de 0,05. Il s'ensuit que l'on peut s'attendre à ce que 5 % des essais de toxicité présentent une différence attribuable au seul hasard, d'où la possibilité de commettre une erreur* **α** *une fois sur 20.*
- **β** *(bêta) est la probabilité de commettre une erreur* **β**. *Elle est inversement proportionnelle à* **α**, *de sorte que, si l'on choisit de faibles valeurs de* p, *la probabilité d'erreur β augmente. La puissance d'un essai* (**1 − β**), *est la capacité de discrimination d'un essai. La plupart des expérimentateurs ne choisissent pas le plan*

Tableau 4. — Types d'erreur dans les tests d'hypothèses et probabilités associées (d'après USEPA et USACE, 1994).

Conclusion tirée du test d'hypothèse	Situation réelle des populations	
	Aucune différence (H_0 est vraie)	Différence (H_0 est fausse)
Acceptation l'hypothèse nulle (on conclut que H_0 est vraie)	Correct (probabilité = 1 - **α**)	Erreur β (probabilité = **β**) « faux négatif »
Rejet de l'hypothèse nulle (on conclut que H_0 est fausse)	Erreur α (probabilité = niveau de signification = **α**) « faux positif »	Correct (probabilité = puissance = 1 – **β**)

d'expérience en fonction de **β** *ni de la puissance, bien que cela puisse être souhaitable.*

- *Appliquée à une situation réelle, l'erreur α aboutirait, en écotoxicologie, à des limites de rejet plus rigoureuses ou à un traitement supplémentaire des déchets, mesures qui pourraient ne pas être nécessaires. L'erreur β augmenterait la probabilité d'atteinte à l'environnement.*
- *On devrait signaler la* différence significative minimale (DSM) *comme une autre façon de décrire la puissance d'un essai de toxicité. La DSM est la plus petite différence en pourcentage entre les résultats obtenus avec le témoin et les résultats obtenus avec un traitement qui serait statistiquement différente dans le cadre du plan d'expérience de l'essai de toxicité.*
- *L'application inverse de la DSM est la « bioéquivalence » ; avant d'entreprendre l'essai de toxicité, on fixe, comme critère de validité de l'essai, un degré de différence acceptable entre le traitement et le témoin.*
- *On ne devrait normalement pas appliquer le test d'hypothèse(s) à des données quantiques non transformées. On peut cependant l'appliquer si les répétitions réunissent des observations quantiques sur au moins 100 individus, nombre est suffisamment élevé pour s'approcher d'une distribution quantitative. Il est actuellement quelque peu utilisé, bien que des méthodes supérieures puissent être désormais exigées.*

7.2.1 Répétitions et unités expérimentales

Dans le test d'hypothèse(s), il importe beaucoup d'identifier les unités expérimentales et les vraies répétitions ou échantillons réitérés (explications dans le § 2.5 et avertissement ici). Une désignation des *répétitions* qui manquerait de rigueur pourrait mener à une analyse et à des conclusions extrêmement erronées. En particulier, les organismes se trouvant dans une enceinte expérimentale unique ne seraient pas des répétitions, mais des *unités d'échantillonnage*.

7.2.2 Erreurs α et β

Dans un test d'hypothèse, il est particulièrement facile de formuler des conclusions erronées d'une manière soit excessivement optimiste, soit excessivement prudente. Le sujet est étroitement apparenté à la notion de *puissance* du test statistique utilisée dans le test d'une hypothèse (v. le § 7.2.3). La notion est également reliée à la question de signification statistique par opposition à celle de signification biologique (§ 9.3).

On commet une *erreur α* ou *de première espèce* (« faux positif ») quand on rejette l'hypothèse nulle, alors que cette dernière est en réalité vraie (c'est-à-dire que l'on conclut à une différence qui n'existe pas). On commet une erreur β ou de seconde espèce (« faux négatif ») quand on accepte (c'est-à-dire quand on ne rejette pas) l'hypothèse nulle d'une absence de différence, qui existe en réalité. Le tableau 4 montre les relations existant entre les conclusions des tests et la vraie situation (inconnue).

La plupart des expérimentateurs maîtrisent partiellement ces erreurs en fixant le niveau de signification (***α***) pour la tolérance de faux positifs (erreur α). Presque toujours, **α** est fixé de façon à ce que la probabilité (*p*) égale 0,05. Dans ce cas, on peut s'attendre à ce qu'un test sur 20 (5 % ou 0,05) révèle une différence apparemment significative par le seul effet du hasard, c'est-à-dire que les éléments que l'on compare sont fortement divergents, mais pas vraiment différents. En conséquence, il existe une probabilité de 1/20 de conclure à un « faux positif » ou de commettre une erreur α. Si on attribuait à **α** une valeur plus grande (disons 0,1), la probabilité de conclure à un faux positif augmenterait (dans cet exemple, on s'attendrait à un écart de probabilité dans un essai sur 10). Si **α** était petit (0,01), on s'attendrait à ce que seulement un essai sur 100 donne lieu à une erreur α (mais lire le texte qui suit pour savoir à quel prix).

La probabilité de commettre une erreur β (ou de seconde espèce) est appelée ***β*** (bêta), la probabilité d'accepter l'hypothèse nulle quand, en réalité, elle est fausse (tableau 4). La valeur de **β** est rarement fixée de façon délibérée par l'expérimentateur avant la réalisation de l'essai (v. le texte qui suit), mais elle est déterminée en grande partie par le choix initial d'α. Il existe une relation inversement proportionnelle entre **α** et **β,** et, à mesure que diminue le niveau de signification (plus la valeur attribuée à **α** est petite, moins on risque de commettre l'erreur α), β augmente et, de la sorte, la probabilité de commettre l'erreur β. Autre facteur : plus le plan de l'expérience est puissant (par ex. plus de répétitions, § 7.2.3), moins l'erreur β est probable,

Les statisticiens relient habituellement β à la *puissance* du test, $1 - \beta$, et que l'on peut définir comme suit :

a) la « capacité de discrimination » du test ;
b) la probabilité de conclure correctement qu'il existe une différence ou (mieux encore) ;
c) la « probabilité de rejeter l'hypothèse nulle quand celle-ci est en réalité fausse et devrait être rejetée ».

Lorsque l'on applique les résultats d'essais de toxicité au monde réel, les erreurs α et β ont des conséquences très différentes. Conclure à tort à la présence d'un effet toxique (erreur α), si cela s'applique à un rejet industriel ou à la fixation de limites de qualité de l'eau, pourrait mener à des restrictions plus rigoureuses ou à un traitement plus poussé des déchets. Les conséquences seraient une marge plus large de sécurité pour la nature et un coût accru pour l'activité humaine [52]. D'autre part, ne pas déceler une différence réelle (erreur β) pourrait donner une impression non justifiée de sécurité à l'égard de la matière soumise à l'essai, ce qui aboutirait probablement à un milieu récepteur nocif pour les organismes. Du point de vue écologique, les erreurs β sont plus graves. En conséquence, le niveau de signification (**α**) ne devrait pas être fixé à des niveaux excessivement rigoureux. Le choix d'un niveau de signification de 0,01 au lieu de 0,05 pourrait sembler une norme rigoureuse, mais cela diminuerait aussi la puissance du test, augmenterait la probabilité d'erreur β et augmenterait la possibilité de conséquences nuisibles pour l'environnement.

7.2.3 Puissance d'un essai de toxicité

Dans le § 7.2.2, nous avons présenté la notion de *puissance* d'un test statistique dans le test d'hypothèse(s). La puissance subit l'influence de plusieurs facteurs :

- le niveau de signification (**α**) choisi par l'expérimentateur ;
- la variabilité des répétitions ;
- la grandeur de l'effet (c'est-à-dire de l'effet véritable, visé par l'essai) ;
- *n,* le nombre d'échantillons ou de répétitions utilisées dans l'essai.

Le calcul de la puissance (d'un essai) peut servir a priori à la détermination de la grandeur de l'erreur β et à la probabilité de résultats qui sont des faux négatifs (USEPA et USACE, 1994). L'expérimentateur peut choisir trois des quatre éléments énumérés dans la liste qui précède et les intégrer dans le plan d'expérience. Le quatrième, la variabilité, est difficile à prévoir, mais on peut l'estimer à partir d'expériences antérieures ou d'essais préliminaires. En conséquence, il pourrait être

52. Grâce à une expérience au résultat surprenant, Moore *et al.* (2000) ont montré l'importance du choix du niveau de puissance et du niveau de signification statistique. Ils ont demandé à des laboratoires d'analyser des échantillons d'eau synthétisée en laboratoire, non toxique, en leur faisant croire qu'il s'agissait d'échantillons d'eau usée. 6 des 14 laboratoires ont signalé que l'eau était toxique. Moore *et al.* (2000) n'ont pas pu trouver de raisons plausibles à ce niveau élevé d'erreur α. Ils proposent des solutions, notamment des critères supplémentaires pour l'acceptation des essais de toxicité.

long ou ennuyeux d'intégrer la puissance dans le plan d'expérience d'un essai, ce qui explique pourquoi on ne le fait pas souvent. Un plan d'expérience prévoyant une puissance convenable pourrait signifier la nécessité d'un essai d'envergure, ce qui, économiquement et logistiquement, est peu attrayant. Dans ce cas, l'expérimentateur devrait au moins reconnaître les limites de l'essai et la possibilité d'aboutir à une conclusion erronée.

Aucune valeur standard n'a été mise au point pour la puissance d'un essai, ni pour son fondement, qui est le taux d'erreur β. Pour β, le taux d'erreur de 10 % (la puissance = 90 %) a été adopté pour la surveillance des effets dans les mines de métaux (EC, 2002b), et on pourrait le considérer comme un objectif convenable. Cependant, même à ce taux, toute conclusion sur l'absence d'effet toxique pourrait être peu fiable. À un taux de puissance de 90 %, un essai sur 10 pourrait ne pas révéler d'effet, du seul fait du hasard, peut-être en raison d'un échantillon de petite taille ou de la variabilité des organismes. Il est sage de tempérer les conclusions selon lesquelles « il n'y a pas effet » en ajoutant : « pour ce plan d'expérience et un essai de cette puissance ». Pour un essai de faible puissance, il pourrait être plus réaliste d'annoncer un résultat peu concluant plutôt que d'affirmer qu'il n'y a pas d'effet.

On a enjoint aux toxicologues de préciser **α** et la puissance statistique (1 – β) comme indications de la possibilité de tirer des conclusions erronées dans un sens comme dans l'autre. La plupart des gens éprouvent des difficultés avec la puissance et ne la signalent pas. En effet, les formules précises de cette notion assez complexe diffèrent selon les divers tests statistiques. Vu cette complexité, nous exposons dans le § 7.2.4 une solution de rechange utilisant la *différence significative minimale*. Les expérimentateurs souhaitant signaler la puissance d'expériences de toxicité devraient consulter USEPA et USACE (1994).

7.2.4 Différence significative minimale

La *différence significative minimale* (DSM) est un cas particulier de la puissance d'un essai donné, et on peut la considérer comme un indice de la puissance . Comme la DSM est une caractéristique du logiciel employé pour beaucoup de tests de comparaisons multiples (§ 7.5), le fait de la signaler remédie en partie à la difficulté de communiquer la puissance d'un essai de toxicité.

La signification exacte de *différence significative minimale* dépend du test statistique envisagé. En général, la DSM est la grandeur de la différence qui devrait exister dans les mesures moyennes (poids, par ex.), entre le témoin et une concentration d'essai, pour qu'on puisse conclure à un effet significatif à cette concentration. Manifestement, la DSM augmente en même temps que la variation à chaque concentration.

On exprime souvent la DSM en pourcentage. Par exemple, une DSM de 12 % signifierait qu'une différence de 12 % entre les mesures correspondant à une concentration et celles qui correspondent au témoin serait la différence minimale décelable dans l'essai de toxicité. (Autrement dit, si une différence de 12 % était constatée, on la considérerait comme statistiquement significative, pour le mode opératoire utilisé).

Si on signalait la CSEO et la CEMO comme paramètres de toxicité, il serait avantageux de préciser la DSM. L'utilisateur des résultats aura une idée quelque peu meilleure de la variabilité dans un essai donné et de l'interprétation plus ou moins étroite à donner aux résultats. Environnement Canada exige la déclaration de la DSM dans les rapports assujettis à ses programmes, pour les tests statistiques permettant de la déterminer. Il est fortement recommandé de préciser aussi la DSM dans les rapports relevant d'autres autorités (Miller *et al.*, 1993).

Tous les tests de comparaisons multiples paramétriques tels que les tests de William et de Dunnett (§ 7.5) donnent la DSM ou son équivalent. Malheureusement, dans la pratique courante actuelle, les tests non paramétriques (tels que le test multiunivoque de Steel ou le test de sommation des rangs de Wilcoxon) ne donnent aucun analogue utile.

Valeurs acceptables de la DSM. — Jusqu'ici, Environnement Canada ne s'est donné aucune ligne directrice pour décider de l'acceptabilité d'une DSM.

À la réunion du Groupe consultatif sur la statistique, il a été envisagé, mais non décidé, d'adopter une limite, qui permettrait, par ex., d'invalider un test dont la DSM excéderait 50 % (Miller *et al.*, 1993).

L'État du Washington a adopté une DSM de 40 % pour les essais de toxicité sublétale, pour les besoins de la réglementation (WSDOE, 1998). L'USEPA (2000b) a proposé des maximums recommandés pour l'acceptation des résultats de certains essais de toxicité (tableau 5). Les valeurs découlent de l'examen d'une base nationale de données sur 23 modes opératoires utilisés sur des toxiques de référence pendant une décennie dans 75 laboratoires. Peu importe la valeur de probabilité (**α**) choisie, les maximums s'appliqueraient.

Il semble, à la lecture du tableau 5, que la valeur normale de la DSM est beaucoup plus une caractéristique individuelle des divers essais de toxicité. Apparemment, il ne convient pas d'attribuer une valeur à la DSM pour tous les organismes et tous les modes opératoires. Dans les essais de mesure d'un double effet, les valeurs recommandées ne s'appliquent qu'à l'effet sublétal.

Dans une étude objective, Wang *et al.* (2000) sont arrivés à la même conclusion (nécessité de DSM différentes pour différents essais). Ils ont conclu, à partir d'essais avec des ensembles appropriés de données, que les limites de la DSM pouvaient être fixées scientifiquement, grâce à une équation assez complexe, fournie par eux, qui dépendait de plusieurs autres variables, notamment de la puissance du test statistique et de la différence décelable souhaitée par rapport au témoin. Aucune valeur unique, calculée mécaniquement, n'a pu être donnée.

7.2.5 Bioéquivalence

Bioéquivalence est le nom donné au test relié à la DSM. Cet outil de test d'hypothèse a pour effet d'inverser l'approche générale et d'utiliser l'hypothèse nulle à l'envers. On fixe d'abord un degré de différence acceptable, entre les performances du témoin et celles qui correspondent aux concentrations d'essai. L'hypothèse nulle est la suivante : les résultats de l'essai ne s'écartent pas de la normale de plus de la différence acceptable. On teste cette hypothèse au moyen d'un traitement statistique.

Shukla *et al.* (2000) ont montré les avantages de la bioéquivalence. Beaucoup d'essais de toxicité dans lesquels on a constaté un effet appréciable et que l'on a jugés valides en vertu de la méthode classique, en raison de la forte variabilité interne de l'essai, ont échoué (de façon méritée) en vertu de l'approche dite de bioéquivalence. Beaucoup d'essais n'ayant révélé qu'un effet légèrement toxique, dans lesquels la matière à l'étude avait « échoué » en vertu de la méthode classique, à cause d'une légère variabilité interne de l'essai, ont reçu une note (méritée) de passage en vertu de la méthode dite de bioéquivalence. Les bases statistiques de la méthode de bioéquivalence sont exposées dans Wellek (2002).

La méthode de bioéquivalence exige l'accord sur ce qui constitue un effet biologique significatif, ce qui n'est pas décidé pour la plupart des essais d'Environnement Canada (v. le texte qui précède). Cependant, dans les essais canadiens, on a commencé à vouloir définir les limites d'un effet acceptable. La surveillance réglementaire des effluents industriels exige habituellement que les essais de létalité aiguë montrent un effet de moins de 50 %. Cela ne signifie pas que la mort de près de la moitié des organismes en expérience est acceptable. Ce paramètre de toxicité a été adopté parce qu'on pouvait l'estimer avec une exactitude et une fiabilité convenables. En outre, on pensait qu'une maîtrise suffisante de la toxicité dans le déchet rejeté permettrait la mise en place de conditions satisfaisantes après la dilution dans le milieu récepteur.

Le Programme canadien d'immersion en mer a des exigences quelque peu plus contraignantes pour deux essais de toxicité.

Le sédiment ne passe pas l'essai effectué avec des oursins (EC, 1992f) si la réussite de la fécondation est inférieure de 25 % au taux de réussite dans l'eau témoin. Dans l'essai employant des amphipodes marins (EC, 1992d ; 1998b), même constat si la survie est inférieure de 20 % au taux de survie dans le sédiment de référence ou inférieure de 30 % au taux de survie dans le sédiment témoin (Porebski et Osborne, 1998 ; Zajdlik *et al.*, 2000). Il faut également que la différence soit statistiquement significative. Autrement dit, l'effet apparemment nuisible ne devrait pas résulter du hasard. Il faut également satisfaire aux critères de validité d'un essai de toxicité. Les scientifiques du Programme d'immersion en mer ont voulu que ces limites soient raisonnablement représentatives d'une différence écologiquement significative par rapport à la variabilité naturelle des populations. Ils étaient grandement conscients des connaissances limitées ayant présidé à la

Tableau 5. — Différences significatives minimales (DSM) recommandées par l'USEPA pour des effets sublétaux manifestés dans certains essais de toxicité (d'après USEPA, 2000b).

Méthode d'essai publiée par l'USEPA	Effet mesuré	DSM maximale
Ceriodaphnia, reproduction et survie	Reproduction	37 %
Tête-de-boule, survie et croissance des larves	Croissance	35 %
Capucette béryl (*Menidia beryllina*), survie et croissance des larves	Croissance	35 %
Crustacé mysidé, survie, croissance et fécondité	Croissance	32 %
Cyprinodon varié (*Cyprinodon variegatus*), survie et croissance des larves	Croissance	23 %
Pseudokirchneriella subcapitata, croissance et reproduction	Croissance et reproduction	23 %

fixation de ces limites, mais, manifestement, des limites étaient nécessaires pour les programmes réglementaires. La validation du choix est le sujet d'une étude (Zajdlik *et al.*, 2000).

7.2.6 Emploi des techniques sur les données quantiques

On peut utiliser le test d'hypothèse(s), technique normalement quantitative, pour évaluer les effets quantiques, mais, d'habitude, on ne devrait pas le faire. Une exception, cependant, serait si les données étaient convenablement transformées, comme il est exposé dans les § 2.9.2 et 2.9.3.

Une autre exception serait si chaque répétition renfermait au moins 100 observations ; l'analyse quantitative pourrait servir directement, comme nous en avons discuté dans le § 6.1.1. Un exemple est l'essai de fécondation d'oursins (EC, 1992f) dans lequel, avec des données quantiques sur la fécondation de 100-200 œufs par récipient, il est satisfaisant de traiter les données comme si elles représentaient une distribution continue.

Si les répétitions comptent peu d'individus, disons moins de 100, les résultats doivent être analysés comme s'ils étaient des données quantiques.

7.3 Préparatifs du test par analyse de variance

Le test d'hypothèse(s) est une méthode bien reconnue, possédant une approche générale, communément utilisée dans la recherche en pharmacologie et sur la santé humaine. Il dispose d'un ensemble de techniques statistiques, utilisables sur des données quantitatives, c'est-à-dire variables en grandeur entre les individus, par ex. la taille, le poids ou le nombre de tumeurs [53].

Repères

- *Pour le test d'hypothèse(s) appliqué à des résultats quantitatifs, on recherche des différences statistiques dans les effets de l'exposition à différents traitements. Il doit y avoir des répétitions. Souvent, les différents traitements seraient une série de concentrations et un témoin (le texte qui suit répond à cette hypothèse).*
- *Le test de Shapiro-Wilk sert à estimer la normalité de la distribution des données, tandis que le test d'O'Brien (ou celui de Levene ou de Bartlett) permet de juger de l'homogénéité des variances des divers traitements. Grâce à des*

53. Certains essais dont les résultats sont quantiques peuvent être analysés au moyen d'un test d'hypothèse(s) si le nombre d'observations est grand (§ 7.2.6).

résultats favorables, l'expérimentateur peut passer aux analyses paramétriques. Le traçage d'un graphique pourrait aider à estimer la normalité.

- *Si les données ne se conforment pas à la normalité ni à l'homogénéité de la variance, on pourrait les y amener en leur faisant subir une transformation. L'analyse pourrait ensuite se poursuivre avec les méthodes paramétriques classiques.*
- *Si les données transformées ne sont toujours pas conformes à la normalité et à l'homogénéité de la variance, il faut employer des méthodes non paramétriques d'analyse. Les analyses paramétriques ne seraient pas valides, mais on pourrait également en réaliser pour comparer les sensibilités estimées. Les méthodes paramétriques sont relativement robustes pour les petits écarts par rapport à la normalité et à l'homogénéité de la variance ; pour les écarts légers, les résultats de l'analyse paramétrique pourraient être communiqués en sus de ceux, qui sont exigés, de l'analyse non paramétrique.*

La démarche fondamentale consiste à adopter une *hypothèse nulle* selon laquelle les effets manifestés par les organismes exposés aux concentrations ne seront pas différents de ceux que l'on constate chez les organismes témoins. On effectue ensuite l'essai de toxicité et on mesure les degrés d'effet dans les groupes répétés d'organismes exposés à une ou à plusieurs concentrations ainsi que chez les organismes témoins. Quand on utilise des méthodes paramétriques, la comparaison des degrés d'effet révèle si les différences entre les groupes différemment exposés (*variation intergroupes*) sont statistiquement supérieures à la *variation intragroupe* (correspondant à chaque traitement) d'ensemble. Dans les méthodes non paramétriques, la comparaison se fonde sur le classement relatif des traitements. Si le traitement n'a aucun effet, le classement moyen devrait être le même pour les divers traitements. Si on ne décèle aucune différence entre une concentration expérimentale, n'importe laquelle, et le témoin, comparativement au « bruit » général entre les répétitions, l'expérimentateur accepte l'hypothèse nulle, c'est-à-dire que la ou les conditions expérimentales ont eu un effet nul. S'il existe une ou des différences significatives entre le ou les traitements, d'une part, et le témoin, d'autre part, on rejette l'hypothèse nulle et, automatiquement, on accepte l'*hypothèse alternative,* selon laquelle la matière à l'étude a exercé un effet réel, c'est-à-dire de la toxicité.

Les méthodes statistiques générales employées en écotoxicologie ont connu un bon développement dans les années 1980 et 1990 (travaux cités dans l'annexe P). Wellek (2002) donne les renseignements de base statistiques. La fig. 19 présente l'organigramme général du test d'hypothèses. D'habitude, si l'essai de toxicité est bien planifié et produit des effets constants chez les organismes en expérience, il suivra un parcours vertical vers le bas, dans le centre de cette figure, tandis que les essais souffrant de quelque irrégularité ou de quelque problème se déporteront vers la droite. Les essais de comparaisons multiples les plus recommandés sont indiqués au bas de la figure, avec des solutions de rechange si le premier choix n'est pas accessible. Quelques autres sont mentionnées dans le texte.

L'expérimentateur devrait représenter les résultats de l'essai sur un graphique, même si test d'hypothèse(s) ne permet pas d'ajuster une droite aux données. L'examen du tracé à permet d'évaluer le caractère convenable de la CSEO et de la CEMO et de constater toute anomalie dans les données (v. les exemples du § 10.4).

Bien que les concentrations expérimentales aient dû être choisies dans une suite géométrique (§ 2.2), dans les circonstances habituelles, l'échelle de concentration n'est pas un facteur dans l'analyse statistique, laquelle porte sur les effets. Les concentrations ne servent que d'étiquettes aux groupes.

7.3.1 Tests de normalité et de comparaison de variances

L'analyse de variance est au cœur du test d'hypothèse(s) dans l'analyse paramétrique. Elle se fonde sur des hypothèses selon lesquelles les données obéissent à la distribution normale et que les variances des divers groupes ou traitements sont semblables. Les mêmes hypothèses s'appliquent aux tests paramétriques de comparaisons multiples qui suivent l'analyse de variance. L'expérimentateur doit vérifier s'il est satisfait à ces hypothèses avant d'appliquer l'analyse de variance. Les tests sont énumérés dans le présent

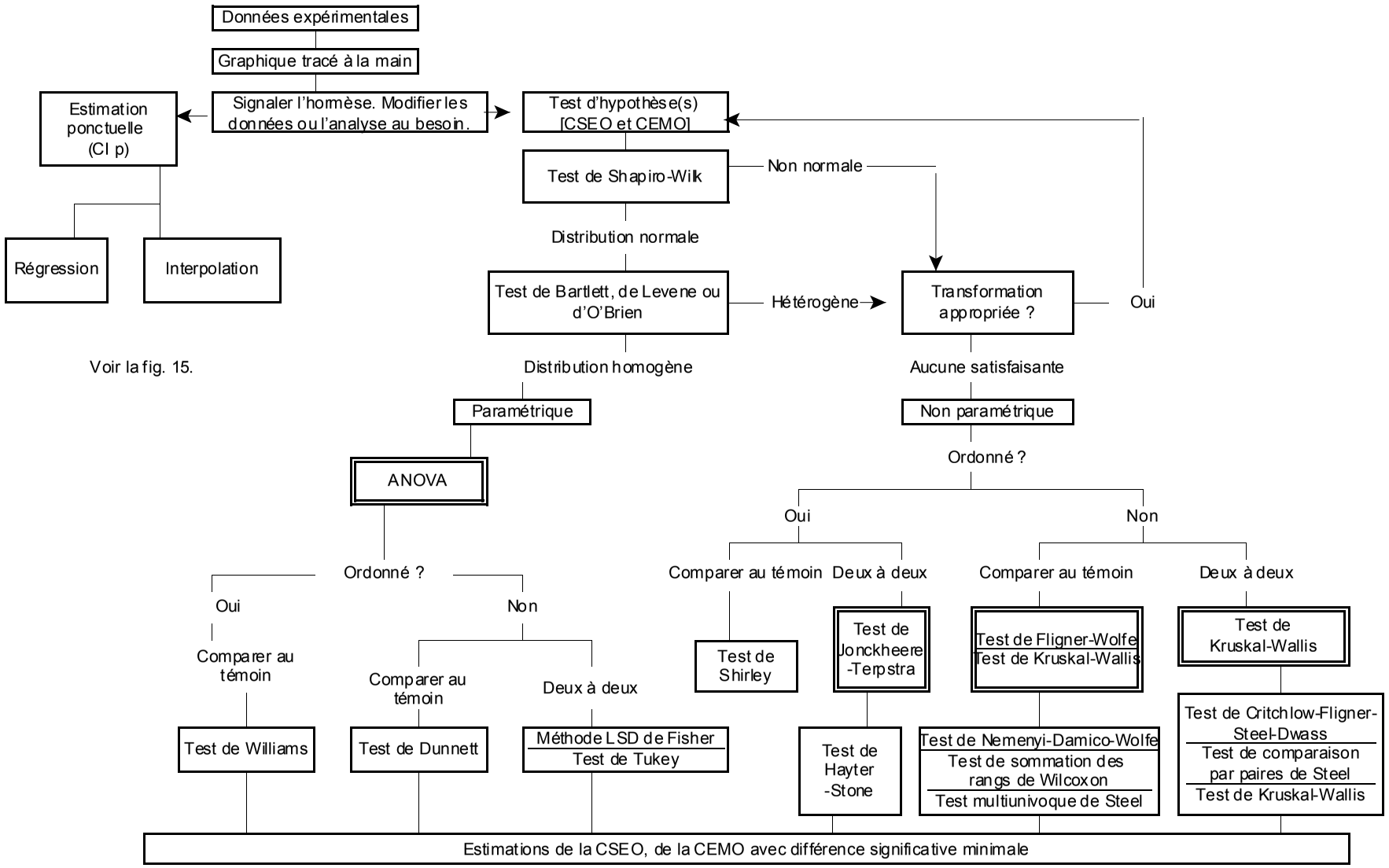


Figure 19. — **Organigramme des analyses statistiques pour les tests d'hypothèses dans les essais de toxicité.** Les cases entourées d'une double bordure dénotent des tests d'hypothèse nulle. Ce n'est que si cette hypothèse est rejetée que l'analyse peut passer au test de comparaisons multiples.

paragraphe et sont décrits plus en détail dans le § P.2 de l'annexe P. Il doit y avoir au moins deux répétitions pour tous ces tests statistiques et il est souhaitable qu'il y en ait davantage ; les déficiences des tests de comparaisons multiples s'aggravent si le nombre de répétitions diminue.

Si l'un des tests de qualification échoue, il faut analyser les données par des méthodes non paramétriques de rechange (§ 7.5.2). Si la non-conformité est petite, il pourrait être avantageux d'effectuer des analyses paramétriques et non paramétriques et d'en communiquer les résultats (§ 7.3.2).

Normalité. — On recommande le test de Shapiro-Wilk pour tester la normalité, plutôt que le test de Kolmogorov-Smirnov, offert dans certains programmes informatiques. Le test de Shapiro-Wilk est décrit dans le § P.2.1 de l'annexe P, accompagné d'un exemple. L'analyse se fonde sur les résidus, avec un échantillon dont la taille minimale est de 3. Les programmes informatiques usuels de toxicologie s'occupent des calculs compliqués. La comparaison finale utilise une valeur critique (*W*), trouvée dans les tables (Shapiro et Wilk, 1965 ; D'Agostino, 1986), et l'expérimentateur peut évaluer le degré de non-conformité.

En outre, l'obtention d'un graphique des données pour chaque répétition ou concentration pourrait être instructive (v. la fig. P.1, dans l'annexe P). Le graphique pourrait laisser entrevoir la cause apparente de la non-normalité ou de la non-homogénéité.

Homogénéité de la variance. — Nous recommandons le *test de Levene* (1960), mais, malheureusement, ce test ne fait pas partie des progiciels conçus pour l'écotoxicologie. Le test de Levene évite un problème qui se manifeste dans le test de Bartlett : l'extrême sensibilité aux données non normales. Le test de Levene se fonde sur la moyenne des écarts absolus des observations par rapport à la moyenne des traitements. Il est peu facile à trouver, mais on pourrait l'appliquer par traitement manuel des données (§ P.2.3, annexe P).

Le *test d'O'Brien* (O'Brien, 1979) est supérieur à celui de Levene sous certains aspects mathématiques ; cependant, il est presque introuvable, même dans les manuels.

Le *test de Bartlett* (1937) est usuel dans les progiciels d'écotoxicologie pour tester l'homogénéité de la variance. Nous le décrivons dans le § P.2.3 de l'annexe P. Il a l'inconvénient d'être très sensible aux données qui ne pas suivent la loi normale, particulièrement les distributions asymétriques. Un ensemble de données pourrait être rejeté à cause d'une conclusion erronée sur l'homogénéité de la variance.

Le point de chacun de ces tests est l'hypothèse nulle selon laquelle il n'existe pas de différence entre les variances des traitements. Si les variances diffèrent de façon importante, une analyse de variance ultérieure est invalide. Ces tests reposent sur l'hypothèse selon laquelle les observations suivent la loi normale. Les données fondées sur des proportions ne devraient normalement pas être soumises à ces tests (§ P.2.4 de l'annexe P).

7.3.2 Décisions après le test de distribution données

Les résultats qui réussissent les tests de Shapiro-Wilk et de Levene ou de Bartlett devraient être analysés par des méthodes paramétriques, c'est-à-dire l'analyse de variance. Les données qui ne satisfont à aucun de ces tests pourraient être transformées pour répondre aux exigences. On soumet les données transformées à des tests de normalité et d'homogénéité et si elles satisfont aux exigences, on les analyse à l'aide de méthodes paramétriques usuelles. La transformation entraîne cependant des complications et des inconvénients, décrits dans le § 2.9.2.

Si les données originales ou transformées ne satisfont à aucun test de distribution des données, l'analyse doit employer des méthodes non paramétriques (fig. 19).

Les tests de normalité et d'homogénéité de la variance peuvent être excessivement sensibles, parfois, tandis que les tests d'analyse de variance et de comparaisons multiples sont plutôt robustes à l'égard des non-conformités mineures (§ P.2.4). En conséquence, si un ensemble de données déviait légèrement ou modérément de la normalité ou de l'homogénéité de la variance, l'expérimentateur pourrait souhaiter consulter un statisticien sur les méthodes convenables d'analyse. Nous recommandons la réalisation d'une analyse paramétrique et non paramétrique et d'en communiquer les résultats. La plus sensible des deux analyses devrait

donner l'estimation définitive de la toxicité [54].

On devrait également présenter les résultats des tests de Shapiro-Wilk's et d'O'Brien (ou de Bartlett) avec un graphique des résultats bruts. La raison en est que les tests paramétriques sont souvent plus puissants pour déceler des effets toxiques, même quand les données renferment des irrégularités mineures.

7.4 Analyse de variance

On effectue une *analyse de variance* pour les tests paramétriques. On teste ainsi *l'hypothèse nulle* (H_0) selon laquelle il n'existe pas de différence dans l'effet moyen entre les traitements (concentrations). La plupart des expérimentateurs connaissent bien l'analyse de variance, et cette dernière est offerte dans la plupart des progiciels de toxicologie. Elle est davantage décrite dans le § P.3. Dans l'analyse de variance, on compare la variation entre les effets moyens correspondant aux divers traitements (concentrations) et la variation des effets correspondant aux répétitions de chaque concentration. On compare le quotient entre les deux variations aux valeurs critiques présentées dans les tables pour déterminer s'il existe une ou plusieurs différences significatives entre les traitements. S'il n'existe pas de telles différences, l'analyse se termine là, et on adopte l'hypothèse nulle. Si, au contraire, l'analyse révèle une différence, elle peut se poursuivre par un test de comparaisons multiples pour identifier les différences.

54. Il semble y avoir, à cet égard, des appuis pour plus de souplesse. Un groupe de statisticiens et d'autres professionnels ayant rédigé une analyse sur l'écotoxicité pour l'OCDE (OECD, 2004) avait une opinion étonnamment détendue sur les tests formels de normalité et d'homogénéité. Dans le paragraphe consacré au choix entre des méthodes paramétriques et non paramétriques, il déclare que « l'inspection visuelle des données peut avoir montré que la dispersion est plus ou moins symétrique et homogène... que, dans ce cas, on peut vouloir analyser les données par les méthodes paramétriques usuelles, fondées sur la normalité ». Plus loin, le groupe ajoute que, « lorsque les données semblent se conformer aux hypothèses (après inspection visuelle) d'une analyse paramétrique particulière, c'est la méthode évidente à choisir. On peut vérifier plus avant les hypothèses à la fin de l'analyse (par ex. en examinant les résidus ...). On peut noter que l'analyse paramétrique fondée sur des hypothèses normales est raisonnablement robuste aux légères infractions contre les hypothèses. » « Des tests formels existent également..., mais il est à noter qu'une infraction légère aux hypothèses n'est pas préoccupante, et les tests ne mesurent pas le degré d'infraction. Pour la plupart des expérimentateurs, il serait difficile de juger de ce qui constitue une « infraction légère », auquel cas on devrait obtenir les conseils d'un statisticien.

Repères

- *Lorsque les données sont conformes à la normalité et à l'homogénéité de la variance, la première étape du test paramétrique est une analyse de variance pour déceler une différence globale entre les traitements. Dans l'analyse de variance, on compare la variation entre les concentrations et la variation de fond correspondant à chaque concentration.*
- *Si l'analyse de variance permet de déceler une différence globale, on la fait suivre d'un test de comparaisons multiples pour décider quelle(s) concentration(s) a (ont) causé des effets différents de ceux du témoin. Cela permet de déterminer la concentration avec effet minimal observé (CEMO). La concentration immédiatement plus faible est la concentration sans effet observé (CSEO). On recommande le test de Williams s'il y a un ordre de concentrations dans les traitements ou, dans le cas contraire, le test de Dunnett. On recommande le test de la plus petite différence significative de Fisher pour une comparaison deux à deux (chaque traitement avec chacun des autres). Il existe des tests de remplacement.*
- *Pour l'analyse non paramétrique des données ordonnées, on recommande le test de comparaisons multiples de Shirley pour comparer les traitements au témoin, bien que la méthode ne soit pas facilement accessible. Pour la comparaison deux à deux, le test de Jonckheere-Terpstra serait l'analogue non paramétrique de l'analyse de variance. Si ce test aboutit au rejet de l'hypothèse nulle, on devrait le faire suivre du test de Hayter-Stone pour une comparaison multiple par paires (deux à deux) des effets des traitements.*
- *Pour l'analyse non paramétrique des résultats non ordonnés et la comparaison des traitements avec le témoin, on devrait vérifier l'hypothèse nulle au moyen du test de Fligner-Wolfe. Si l'hypothèse est rejetée, on recommande le test des étendues multiples de Nemenyi-Damico-Wolfe. À*

défaut de pouvoir se servir de ce dernier, les solutions de rechange sont le test de sommation des rangs de Wilcoxon et le test multiunivoque de Steel. Pour la comparaison deux à deux, on devrait utiliser le test de Kruskal-Wallis pour l'hypothèse nulle. Si ce dernier est rejeté, on devrait ensuite appliquer le test de comparaisons multiples de Critchlow-Fligner-Steel-Dwass ; au cas ou ce test ne serait pas accessible, nous énumérons des tests de remplacement.

L'analyse de variance donne ce que l'on appelle la *variance de l'erreur* ou le *terme d'erreur résiduelle* pour tout test ultérieur de comparaisons multiples (§ 7.5). Les programmes informatiques modernes d'analyse de variance peuvent traiter les données dont le nombre de répétitions est inégal et produire le terme correct d'erreur résiduelle pour tout test ultérieur de comparaisons multiples.

7.5 Tests de comparaisons multiples

On applique un test de comparaisons multiples pour déterminer quels traitements provoquent des effets significativement différents des effets observés chez le témoin et, si on le souhaite, différents les uns des autres. Les divers tests de comparaisons multiples (parfois appelés tests des étendues multiples) permettent des comparaisons quelque peu différentes. L'expérimentateur choisit le test approprié (fig. 19). Parce que ceci est un test d'hypothèse(s), aucun des tests ne tient compte de la valeur numérique de la concentration, mais deux d'entre eux (les tests de Williams et de Shirley) tiennent compte des effets moyens dans l'ordre de la concentration et trouvent le premier effet qui diffère de celui que l'on observe chez le témoin. En conséquence, quand les données sont ordonnées, comme dans un essai employant une série de concentrations, la préférence va au test de Williams (paramétrique) ou à celui de Shirley (non paramétrique)[55].

55. Les statisticiens pourraient préférer des approches différentes de celles des tests de comparaisons multiples, au moins pour ce concerne les données paramétriques. Ils pourraient choisir de lancer des comparaisons en utilisant les énoncés intégrés dans les modèles linéaires généraux (GLM) et les modèles linéaires généralisés (GLIM) [v. le § 6.5.2.].

Nous discutons des tests de comparaisons multiples dans les paragraphes qui suivent et nous les expliquons plus en profondeur dans les § P.4 et P.5 de l'annexe P. On trouvera des détails mathématiques dans Newman (1995) ou dans les manuels classiques de statistique. Beaucoup de tests importants sont offerts dans divers progiciels.

7.5.1 Tests paramétriques

Nous recommandons fortement le ***test de Williams*** (Williams, 1972) parce qu'il tient compte de l'ordre des concentrations, croissant ou décroissant. Cette qualité convient à la plupart des essais de toxicité. Le test de Williams compare les effets correspondant à chaque concentration avec les effets observés chez le témoin, comme cela est courant dans beaucoup d'essais à plusieurs concentrations. On compare les statistiques de l'essai, dans l'ordre, à la valeur critique. La première statistique à excéder la valeur critique indique une différence significative de cette moyenne par rapport au témoin. D'une puissance statistique supérieure, le test de Williams est notablement plus sensible pour l'estimation d'une CEMO inférieure que les autres tests disponibles (§ P.4.1, annexe P).

Le test de Williams repose sur l'hypothèse selon laquelle les données correspondant aux concentrations suivent la distribution normale et sont homogènes. Il doit aussi y avoir une suite monotone de concentrations, sinon les moyennes devraient être lissées, bien que cela risque de réduire la sensibilité du test. Le lissage pourrait être offert dans les nouveaux logiciels de toxicologie, sinon on peut l'effectuer par calculs manuels. Les statistiques du test sont estimées par l'une des deux formules simples utilisables, selon qu'il se trouve un nombre égal ou inégal d'observations contribuant aux valeurs moyennes. La valeur critique, correspondant au taux souhaité d'erreur α et aux degrés de liberté de l'erreur, est tirée des tables (Williams, 1972) si les données ne sont pas trop « déséquilibrées » selon les critères exposés par cet auteur. On peut trouver les tables correspondant aux cas déséquilibrés dans Hochberg et Tamhane (1987). Le test perd en puissance lorsque les données ne sont pas équilibrées, et l'OCDE (OECD, 2004) invoque des preuves selon lesquelles on ne devrait pas l'appliquer à des résultats fortement déséquilibrés. Le test de Williams, qui est l'objet d'une autre discussion dans le § P.4.1 de l'annexe P, a été examiné en détail par l'OCDE

(OECD, 2004, annexe), tandis que Newman (1995) en a exposé la marche à suivre.

Le test de Dunnett, comme celui de Williams, compare la moyenne de chaque groupe avec le témoin, mais il est moins puissant parce qu'il ne tient pas compte de l'ordre des concentrations (tableau P.3 ; Dunnett, 1955 ; 1964). Si les échantillons n'obéissent à aucun ordre implicite, par ex. divers sédiments soumis simultanément à un essai à une seule concentration, on peut employer le test de Dunnett plutôt que celui de Williams. Le test de Dunnett a préséance sur celui de Williams, dans les programmes informatiques utilisés en écotoxicologie.

La formule de base du test de Dunnett ressemble à celle du test *t* de Student. Les progiciels courants exigent, pour le test de Dunnett, un nombre égal d'observations à chaque traitement. La publication d'une série de modifications, qui autorisaient des nombres inégaux, a abouti à celle de Dunnett et Tamhane (1998). Tant qu'une modification convenable ne sera pas intégrée les logiciels disponibles, les expérimentateurs possédant de telles données pourraient consulter et utiliser la modification publiée ou utiliser le test de Dunn-Sidak décrit dans l'alinéa qui suit.

On pourrait remplacer les tests de Williams ou de Dunnett par le ***test de Dunn-Sidak,*** si le nombre de répétitions n'était pas égal en raison de pertes accidentelles ou d'autres causes. On utilise souvent l'***ajustement de Bonferroni du test*** **t,** mais il est moins puissant que le test de Dunn-Sidak et il ne confère aucun avantage particulier. Les deux tests sont moins puissants que ceux de Williams et de Dunnett pour l'estimation de la CSEO et de la CEMO.

L'expérimentateur pourrait vouloir comparer les différences entre toutes les paires d'emplacements dans une étude effectuée en plusieurs endroits. Nous recommandons la ***méthode LSD de Fisher,*** apparentée au test *t*. Elle permet de maîtriser le taux d'erreur α global (lié à la famille de valeurs) et de se tirer d'affaire avec un nombre inégal de répétitions, mais elle n'est pas répandue dans les progiciels conçus pour la toxicologie (§ P.4.4, annexe P). La méthode LSD est également destinée uniquement à un petit nombre de toutes les comparaisons possibles dans un ensemble de données, et ces comparaisons devraient être spécifiées d'avance (cette restriction trouve une application générale dans d'autres tests de comparaisons multiples). Le ***test de Tukey*** est semblable, il est généralement disponible, il peut s'adapter à des échantillons de tailles inégales, mais il n'est pas très sensible (tableau P.3 en annexe). Le ***test de Student-Newman-Keuls*** (le test S.N.K.) est une autre solution de rechange.

7.5.2 Tests non paramétriques

Les tests non paramétriques sont de puissants outils pour les données qui ne suivent pas la loi normale. Généralement, ils tendent à être moins puissants que les tests paramétriques, si on les applique à des données obéissant à la loi normale, auquel cas ils pourraient ne pas déceler un véritable effet de toxicité. Beaucoup de méthodes non paramétriques usuelles exigent au moins quatre répétitions ; cependant, certaines se contentent de moins (par ex. le test de sommation des rangs de Wilcoxon).

Il est recommandé que les tests non paramétriques suivent la même séquence générale que celle qui utilisée dans les tests paramétriques. D'abord, l'hypothèse nulle de l'absence de différence dans les traitements devrait être testée à l'aide de méthodes analogues à une analyse de variance. Ce n'est que si l'hypothèse nulle est rejetée que l'on devrait passer à des tests de comparaisons multiples.

Analogues de l'analyse de variance. — Le ***test de la somme des rangs de Kruskal-Wallis*** figure parfois dans les progiciels et peut servir comme équivalent non paramétrique d'une analyse de variance (Kruskal et Wallis, 1952 ; test appelé ci-après *test de Kruskal-Wallis*). Le ***test de Fligner-Wolfe*** (Fligner et Wolfe, 1982) permet d'examiner l'hypothèse nulle selon laquelle les moyennes correspondant aux traitements sont égales ; l'hypothèse alternative habituelle est que les moyennes d'un ou de plusieurs traitements diffèrent du témoin.

Le ***test de Jonckheere-Terpstra*** (Jonckheere, 1954) permet de tester aussi l'hypothèse nulle d'égalité des médianes, mais l'hypothèse alternative est que *les traitements sont ordonnés.* Ce test, très puissant, convient aux données qui s'écartent fortement de la normalité et de l'homoscédasticité. Les échantillons de taille inégale ne lui posent aucun problème, mais le fait

de pas prendre en considération le nombre d'individus dans chaque sous-groupe risque également d'être un inconvénient. Malheureusement, la méthode n'étant pas largement accessible sous forme de programme informatique, elle exige des calculs fastidieux à la main. Cependant, on en trouve une version qui traite les petits échantillons dans les logiciels commerciaux SAS et StatXact (OECD, 2004). Les caractéristiques du test sont décrites en détail dans une annexe de l'OCDE (OECD, 2004).

Les trois tests dont nous venons de parler débutent par l'hypothèse nulle d'égalité de l'effet des traitements. À l'instar des tests paramétriques, si l'hypothèse nulle d'égalité est acceptée, l'analyse statistique s'arrête là, et on conclut à l'absence de différences significatives.

Comparaison multiple. — On recommande le test de Shirley (Shirley, 1977) comme premier choix pour comparer les médianes des traitements avec la médiane du témoin, *s'il existe un ordre dans l'amplitude du traitement et (ou) de ses effets.* Ce test est un analogue non paramétrique du test de Williams et il tient compte de l'ordre des concentrations. Il exige cinq répétitions, mais ces dernières n'ont pas besoin d'être égales. Malheureusement, on ne trouve pas ce test dans la plupart des programmes informatiques, et il n'est pas facile à trouver dans les publications (§ P.5, annexe P).

On peut aussi faire une comparaison deux à deux (chaque traitement avec un autre traitement) si ces traitements sont ordonnés (par ex. une série de concentrations). On peut appliquer le test de Jonckheere-Terpstra, et si l'hypothèse nulle est rejetée, on poursuit l'analyse par le ***test de Hayter-Stone*** de comparaisons multiples deux à deux (Hayter et Stone, 1991). Malheureusement, comme pour les autres tests, le logiciel n'est pas facile à trouver.

Si les traitements ne sont pas ordonnés (par ex. les emplacements dans une étude générale), on devrait appliquer d'abord un analogue non paramétrique de l'analyse de variance. Ce n'est que si l'hypothèse nulle est rejetée (c'est-à-dire qu'il existe une différence quelque part entre les traitements) que l'analyse devrait employer des tests de comparaisons multiples non paramétriques, comme il est décrit dans le texte qui suit. L'étape du test de l'hypothèse nulle n'est pas nécessairement stipulée dans les méthodes exposées ailleurs, mais nous la recommandons dans un souci de prudence. La méthode devrait permettre d'éliminer ou de réduire considérablement les erreurs α, qui consistent à conclure, à tort, à l'existence d'une différence. En termes de statistique, le test de comparaisons multiples est « protégé » par le test initial avec un analogue de l'analyse de variance. Comme on n'effectue pas de test de comparaisons multiples à moins que le test antérieur n'ait rejeté l'hypothèse nulle, le test de comparaisons multiples est « protégé » contre la conclusion de l'existence d'une différence imputable uniquement au hasard.

Si les traitements ne sont pas ordonnés, on recommande le ***test de Fligner-Wolfe*** pour tester l'hypothèse nulle de l'absence de différence d'avec le témoin (Fligner et Wolfe, 1982 ; v. l'annexe P). Si ce test n'est pas offert par un logiciel approprié, on pourrait utiliser le test de Kruskal-Wallis. Si l'hypothèse nulle est rejetée, le premier choix recommandé pour la comparaison avec le témoin est le ***test de Nemenyi-Damico-Wolfe*** (Damico et Wolfe, 1987). Ce test convient à un plan d'expérience équilibré (c'est-à-dire nombre égal de répétitions). Le second choix est le ***test de sommation des rangs de Wilcoxon,*** généralement offert et qui peut prendre en charge les répétitions en nombres inégaux. Ce test également est connu sous d'autres noms tels que *test de Wilcoxon pour observations appariées* et, souvent, en Europe, sous ceux de *test de Wilcoxon-Mann-Whitney* ou simplement de *test U* (§ P.5.4 de l'annexe P). On l'utilise souvent sans tester l'hypothèse nulle, mais cette étape est préconisée par Hollander et Wolfe (1999). Un troisième choix, généralement disponible dans les logiciels toxicologiques, est le ***test multiunivoque de Steel*** (Steel, 1959), qui exige un nombre égal de répétitions.

Si on désire appliquer des comparaisons deux à deux à un ensemble non ordonné de données, on devrait tester l'hypothèse nulle par le test de Kruskal-Wallis. Si l'hypothèse est rejetée, le premier choix du test de suivi devrait être celui de ***Critchlow-Fligner-Steel-Dwass,*** également connu sous le nom de *test de Critchlow-Fligner* (Critchlow et Fligner, 1991). Ce test convient aux nombres de répétitions égaux ou inégaux. S'il n'est pas accessible dans un logiciel approprié, on devrait utiliser, pour les données équilibrées, le ***test de comparaison par paires de Steel*** (Steel, 1960) pour les données équilibrées. Il ne faudrait pas confondre ce test

avec le *test multiunivoque de Steel* (Steel, 1959 ; v. le texte qui précède) précédemment mentionné. Dans le cas des ensembles non équilibrés de données, on pourrait suivre une méthode quelque peu inhabituelle. On teste d'abord l'hypothèse nulle avec le test de Kruskal-Wallis, puis, en cas de rejet, on utilise le même test pour des comparaisons multiples, afin de trouver quelles moyennes des traitements diffèrent les unes des autres.

Edwards et Berry (1987) ont mis au point un test de comparaisons multiples que l'on peut utiliser dans toutes les situations, mais, malheureusement, il n'est pas facile à trouver dans un logiciel.

Essais de mesure d'un double effet

Ces essais mesurent deux effets différents : habituellement la mortalité en tant qu'effet *quantique* et un effet sublétal tel que le poids d'organismes ou le nombre de descendants, qui est presque toujours *quantitatif*. Il a été question de ces catégories d'effets (v. les sections 4, 6 et 7), mais, dans les essais de mesure d'un double effet, les difficultés conceptuelles et statistiques découlent du fait que, souvent, les deux effets interagissent. Par exemple, le poids des individus qui meurent pendant l'essai ne compte pas dans l'évaluation parce qu'il est impossible de le connaître. De même, la mort d'un individu pourrait manifestement influer sur le nombre de jeunes qu'il aurait pu avoir engendrés.

Le choix des méthodes d'analyse est en partie déterminé par les aspects « philosophiques » ou biologiques de l'application des résultats au monde réel et, en partie, par les aspects pratiques des essais particuliers de toxicité. Des paragraphes particuliers sont consacrés, ci-dessous, aux effets quantiques et à deux catégories d'effets sublétaux.

8.1 L'effet quantique

Repères

- *Dans les essais de toxicité sublétale ou chronique, on devrait analyser les effets quantiques à l'aide de techniques quantiques.*
- *Pour ce qui concerne la mortalité survenant pendant un essai visant à mesurer un effet quantitatif chronique ou sublétal, on devrait habituellement effectuer une analyse de la mortalité au moyen de méthodes quantiques usuelles telles que la régression probit.*
- *On pourrait estimer la CL 25 au lieu de la CL 50 si on souhaite un paramètre de toxicité quelque peu analogue à la CI 25 sublétale.*
- *Pour les essais quantiques sublétaux tels que la fécondation d'œufs, on devrait estimer une* CE p *par les techniques quantiques habituelles, bien qu'une analyse quantitative soit possible si on possède au moins 100 observations dans chaque répétition.*
- *Dans les essais de mesure d'un double effet portant sur la reproduction, il pourrait être souhaitable d'analyser cet effet en combinaison avec la mortalité à l'aide d'une approche fondée sur la biomasse.*

Habituellement, le volet quantique d'un essai de mesure d'un double effet est la mortalité, que l'on traite parfois de façon relativement directe. L'expérimentateur ne doit pas supposer que, parce qu'un essai de toxicité est chronique, la mortalité devrait être analysée comme une concentration inhibitrice (CI p). La mortalité est un effet quantique et elle devrait être analysée par des techniques quantiques (section 4). On doit continuer de considérer comme quantiques les données rassemblées sur la mortalité, même si elles proviennent des effets cumulés de diverses actions sublétales survenues pendant une exposition chronique.

On pourrait estimer la CL 25, à l'instar du paramètre quantitatif habituel de toxicité qu'est la CI 25, mais, dans ce cas, l'intervalle de confiance serait plus large que pour la CL 50 (fig. 7). Toutefois, on ne peut pas obtenir ce paramètre par extrapolation : il *doit* y avoir un effet effectivement observé d'au moins 25 % afin d'estimer la CL 25. Cependant, la mortalité maximale pourrait être inférieure à 50 %[56].

On peut estimer d'autres paramètres quantiques de la toxicité grâce aux essais de mesure d'un double effet comme la réussite de la fécondation des œufs de salmonidés, que l'on devrait analyser à l'aide de méthodes quantiques (§ 6.1.1). On peut analyser les

56. Comme nous l'avons fait observer dans le § 4.5.3, certains programmes informatiques d'estimation de la CE p n'analysent pas les données à moins qu'il y ait un effet d'au moins 50 %, mécanisme pour empêcher d'estimer la CE 50 à partir de données inadéquates. Pour la CE 25, il faudrait contourner cette restriction ou, sinon, utiliser une autre méthode.

observations quantiques nombreuses (d'au moins 100 par répétition) par des moyens quantitatifs (§ 6.1.1).

Parfois la mortalité est intimement liée aux effets sur la reproduction, et il convient d'analyser les effets combinés, au moyen d'une méthode fondée sur la biomasse (v. le § 8.3).

8.2 La « croissance » en tant qu'effet sublétal

Repères

- *Dans un essai de mesure d'un double effet (y compris de la taille atteinte* [*la dite « croissance »*]*), il est souvent préférable d'analyser cet effet sublétal séparément de la mortalité, pour estimer un paramètre de la toxicité indépendant, habituellement la CI p. Pour le poids atteint par les alevins de salmonidés, les larves de têtes-de-boule, les jeunes amphipodes ou les larves de chironomes, l'analyse séparée peut se fonder sur le poids final moyen des survivants de chaque répétition. Les individus morts ne fournissent aucune donnée pour l'estimation du paramètre de toxicité fondé sur le poids. Néanmoins, il pourrait y avoir un biais causé par l'interaction, si, disons, des individus « faibles » présentaient à la fois une petite taille et une mortalité rapide ; aucune méthode de résolution de ce problème n'est évidente.*
- *Une solution de rechange, l'approche fondée sur la biomasse, combine mortalité et taille dans l'analyse du poids total des survivants ou du quotient du poids total dans une répétition divisé par le nombre d'organismes au début de l'essai dans cette répétition. On peut utiliser cette approche si on le souhaite ou si elle est prescrite. Dans une certaine mesure, elle simule la réussite écologique et elle pourrait produire des effets plus puissants. Cette approche comporte également le biais éventuel dû à l'interaction de la taille et du temps de survie.*
- *On ne devrait pas, dans un souci de compromis, utiliser en même temps l'approche dite séparée et celle qui est fondée sur la biomasse.*
- *Il faut choisir les techniques mathématiques avec soin. La méthode séparée pourrait entraîner des répétitions dont les effectifs sont déséquilibrés, ce qui limite l'éventail des méthodes statistiques convenables. L'approche axée sur la biomasse pourrait donner des mesures de zéro dans certaines répétitions, ce qui mènerait à des complications avec la variance.*

L'effet sublétal mesuré dans les essais de mesure d'un double effet est d'habitude quantitatif, par ex. le poids ou le nombre de jeunes. L'analyse est plus directe pour ce qui concerne le poids d'organismes, et les choix sont expliqués dans le présent paragraphe (8.2). On qualifie souvent ces essais d'essais « de croissance », mais il serait plus judicieux de parler de « poids atteint » ou de « taille atteinte ». Habituellement, on mesure la taille à la fin de l'essai, mais non au début, comme cela serait exigé pour une bonne évaluation de la croissance.

Le choix d'une méthode pour ce qui concerne le nombre de descendants est plus complexe (v. le § 8.3).

8.2.1 Options de mesure

Dans les essais de mesure d'un double effet portant sur le poids atteint (ou la longueur atteinte ou d'autres mesures de la taille), on devrait prendre sérieusement en considération l'*effet* sublétal que l'on veut analyser et signaler. La valeur du paramètre de toxicité que l'on dégagerait pourrait être beaucoup plus grande ou plus petite, pour certains effets ou certaines combinaisons d'effets. Habituellement, le choix se réduit fondamentalement à combiner des mesures d'un effet sublétal avec la mortalité ou à tenter de les maintenir séparées. Pour les essais de mesure d'un double effet, certaines méthodes d'Environnement Canada précisent la marche à suivre, et, pour les programmes de ce ministère, il faut se conformer à la spécification. Si le choix était laissé à la discrétion de l'expérimentateur, ce serait en partie pour des motifs de doctrine, la ligne de pensée de l'expérimentateur et les applications écologiques des résultats. Néanmoins, le choix a des conséquences certaines sur la validité des méthodes mathématiques.

Quel que soit le choix posé, il pourrait y avoir des interactions inconnues et indésirables subtiles dans les enceintes expérimentales. Par exemple, la mort de certains organismes augmenterait la quantité de

solution, de nourriture et (ou) l'espace dont disposeraient les survivants, ce qui influerait probablement sur leur croissance ou leur mieux-être. Dans l'interprétation des résultats, il faudrait tenir compte de ces possibilités. Bien que l'on n'ait pu faire aucune correction statistique de ces interactions, on peut en réduire l'importance au minimum si on suit les recommandations d'Environnement Canada concernant le volume de la solution d'essai et d'autres questions touchant le mode opératoire.

Une autre difficulté susceptible de conduire à un essai trop sensible de toxicité sublétale a été appelée *interaction entre les maigrichons et les dodus* par le statisticien B. Zajdlik. Il serait tout à fait possible que les individus faibles ou affaiblis, qui seraient de plus petite taille, meurent au cours de l'essai. À une faible concentration où la mortalité serait nulle, ces individus maigrichons survivraient et feraient en sorte que, à cette concentration, le poids moyen, qui serait représentatif, serait relativement faible. À une concentration supérieure, seuls les individus « dodus » survivraient, ce qui, à cette concentration, pousserait le poids moyen vers le haut. L'effet net serait que le paramètre estimé de la toxicité pour l'effet sur le poids serait déplacé vers le bas.

Pour mesurer les effets, trois options se présentent. Les choix sont exemplifiés, dans les lignes qui suivent, par l'essai de toxicité sublétale employant des têtes-de-boule ; cet essai mesure le poids final atteint par un groupe de larves dans une répétition donnée (EC, 1992b). Les options représentent différents objectifs ; elles ne sont pas égales des points de vue biologique ou statistique. Toutes entraînent des difficultés ou possèdent des imperfections mineures ou majeures.

Option 1. — Séparer l'effet sublétal de la mortalité dans l'essai et l'analyser séparément, autant que c'est possible. Cela signifie de dresser un tableau des mesures de l'effet sublétal, uniquement chez les organismes qui ont survécu jusqu'au moment de la prise de mesures (à la fin de l'essai). Dans le cas des larves de tête-de-boule, les données brutes seraient le **poids moyen des poissons survivants.** On diviserait le poids total mesuré dans chaque répétition à la fin de l'essai par le *nombre de larves ayant survécu* dans la répétition. Comme nous l'avons déjà souligné, cela pourrait mener à une « interaction entre les maigrichons et les dodus » d'une ampleur inconnue, contre laquelle il n'y aurait aucun remède. Si aucun poisson ne survivait dans une répétition, il n'y aurait aucune mesure du poids et aucune saisie de données (essentiellement, ce serait une répétition manquante). On évaluerait la mortalité au moyen d'une analyse séparée (§ 8.1).

Option 2. — La prise en compte partielle de la mortalité a parfois été utilisée de la façon inconséquente qui suit et qui n'est pas recommandée. Si, dans une répétition, il se trouvait un ou plusieurs organismes vivants, on estimerait le poids par le poids moyen des poissons survivants, comme dans l'option 1. Si, dans une répétition donnée, les 10 organismes mouraient, on saisirait 0 comme mesure de l'effet sublétal. De la sorte, 0 est le poids moyen des 10 larves mortes dans la répétition, ce qui est absurde. Si toutes les larves étaient mortes, on utiliserait des poids nuls pour les représenter. Si, dans une répétition, des larves survivaient, on ne les représenterait pas par des poids nuls.

Option 3. — Le paramètre de toxicité fondé sur la biomasse résulte de la combinaison d'un effet sublétal et de la mortalité. Il peut entraîner des différences majeures entre les observations à différentes concentrations et insister fortement sur l'effet de la matière à l'étude. Dans l'essai employant des têtes-de-boule, la mesure analysée serait le poids total des poissons vivants, dans une répétition à la fin de l'essai, divisé par le *nombre initial de larves* dans la répétition (Si chaque répétition avait débuté avec le même nombre de larves, on obtiendrait exactement le même résultat par l'analyse du poids total de poissons dans chaque répétition, plutôt que de la moyenne[57].) La **biomasse finale** est la mesure analysée. Si, dans une répétition, tous les poissons meurent, on attribue à cette répétition une masse nulle, comme dans l'option 2 [58].

57. Si, pendant l'exposition, des larves avaient été accidentellement détruites ou perdues, on en soustrairait le nombre du nombre initial de larves dans la répétition.

58. En effet, en vertu de l'option 3, on attribue une masse nulle aux larves mortes dans toute répétition, que la mortalité dans cette dernière ait été totale ou partielle. La méthode est donc une extension de l'option 2, épurée cependant de ses incohérences. La méthode correspond à son appellation d'*approche fondée sur la biomasse.*

Cette approche pourrait aussi être entachée d'un biais : celui de « l'interaction entre les maigrichons et les dodus ».

Les trois options ont été utilisées au Canada, et deux d'entre elles ont été recommandées ou, du moins, conseillées, dans des méthodes publiées par Environnement Canada.

L'option 1 est la pratique courante dans la méthode d'Environnement Canada de détermination du poids des larves de têtes-de-boule (EC, 1992b) [59]. C'est aussi la pratique courante pour l'analyse du poids des alevins de salmonidés dans l'essai au premier stade du cycle évolutif (EC, 1998a), des larves de chironomes (EC, 1997a), de l'amphipode d'eau douce *Hyalella azteca* (EC, 1997b) et du ver polychète *Polydora cornuta* (EC, 2001a) dans les essais de toxicité d'un sédiment.

L'option 2 ne semble être la pratique courante dans aucune méthode publiée. Néanmoins, elle était communément utilisée pour déterminer le poids des larves de chironomes dans les essais de toxicité appliqués par certains consultants canadiens, avant la publication de la méthode par Environnement Canada (1997a).

L'option 3 est courante dans certaines méthodes d'essai de l'USEPA, dans lesquelles le poids ou la taille permet de mesurer l'effet. Dans le programme ICPIN (Norberg-King, 1993), une consigne impose la division du poids total des larves de tête-de-boule dans une répétition par le nombre de larves au début de l'essai.

59. Dans la méthode d'Environnement Canada pour la détermination des effets sublétaux subis par des têtes-de-boule, la consigne est d'exclure une concentration de l'analyse si, dans toutes les répétitions correspondant à cette concentration, toutes les larves sont mortes (EC, 1992b). Les consignes ne sont pas explicites sur la conduite à tenir si toutes les larves sont mortes dans une répétition, mais non dans les autres correspondant à la même concentration. Logiquement, l'analyse ne devrait pas tenir compte de la répétition où la mortalité a été totale. On aurait ainsi un ensemble déséquilibré de données, ce qui exigerait une méthode d'analyse statistique appropriée aux répétitions déséquilibrées. Certes, les consignes d'Environnement Canada n'exigent pas l'application de l'option 2, ce qui aurait signifié la saisie d'une masse nulle pour une répétition dans laquelle toutes larves seraient mortes.

8.2.2 Aspects conceptuels des options

Les trois options du § 8.2.1 présentent divers qualités et défauts.

L'option 1 peut certes être justifiée sur le plan biologique, car elle examine directement le comportement des organismes en expérience ayant subi une exposition complète à la toxicité sublétale et *uniquement* le comportement de ces organismes à la toxicité sublétale. Cette approche semble rationnelle, mais elle peut conduire à des anomalies. Par exemple, dans certains essais de longue durée, avec des amphipodes, la mortalité est un paramètre plus sensible de la toxicité que la croissance. Les sédiments à l'étude pourraient même avoir de meilleures qualités nutritives que le sédiment témoin et favoriser une meilleure croissance des amphipodes (U. Borgmann, Institut national de recherche sur les eaux, Environnement Canada, Burlington [Ont.], communication personnelle, 2001). On remédie à ce type d'anomalie dans les méthodes d'essai d'Environnement Canada, qui exigent une analyse séparée de la mortalité, l'effet le plus sensible étant adopté pour représenter l'essai. En outre, le biais dû à une « interaction entre les maigrichons et les dodus », dont l'ampleur est inconnue, pourrait agir.

L'option 2 est un compromis, historique, que l'on ne peut pas justifier du point de vue conceptuel. Comme nous l'avons mentionné, elle a été utilisée de façon officieuse au Canada pour les essais employant des larves de chironomes dans un sédiment. Visiblement, chaque larve possédait un poids fini au début de l'essai, et le fait de lui attribuer un poids final nul n'a rien de rationnel. À titre d'exemple extrême, si toutes les larves d'une répétition mouraient, la saisie d'une valeur nulle pour leur poids signifieraient qu'il y avait des larves vivantes à la fin de l'essai, mais que leur poids était absolument nul. Cela influe certainement sur la distribution des mesures et abaisse à une valeur inférieure la concentration estimée comme paramètre de toxicité ; cependant l'approche est inconséquence et contre-indiquée.

L'option 3, fondée sur la biomasse, peut se justifier, écologiquement, parce qu'elle simule la réussite globale de l'espèce dans les conditions d'exposition. La réussite écologique se mesure souvent par la biomasse totale ou le nombre total d'individus. L'option 3 est susceptible de donner une courbe dose-effet plus raide que la

courbe résultant de l'option 1, probablement avec une concentration inférieure comme paramètre de toxicité. Cependant, les données de l'option 3 sont plus variables, elles possèdent une sensibilité statistique moins grande pour contrebalancer l'effet biologique apparemment accru (Zaleski et al., 1997). En effet, dans les essais avec les têtes-de-boule, l'option 3 a abouti à des estimations de la toxicité qui sont inférieures à celles de l'option 1 (Pickering et al., 1996 ; WSDOE, 1998). L'option 3 pourrait convenir aux essais à long terme tels que ceux que l'on applique aux sédiments, avec des amphipodes, où la mortalité est le paramètre sensible de toxicité. Encore une fois, cette option pourrait être entachée du biais attribuable à l'« interaction entre maigrichons et dodus ».

8.2.3 Aspects statistiques des options

Les trois options peuvent présenter des subtilités statistiques. Dans chacune, le nombres d'individus pourraient être différent dans les diverses répétitions et aux diverses concentrations, ce qui exige des méthodes statistiques plus complexes. Des nombres non équilibrés ne poseraient pas un problème insoluble aux estimations ponctuelles fondées sur la régression.

L'option 1 semble réserver les problèmes les moins graves à l'analyse. Le nombre inégal d'individus dans les répétitions pourrait être compensé par les méthodes courantes employées en régression ou en analyse de variance. Des difficultés pourraient surgir. Si, à de fortes concentrations, la mortalité devait être générale, ces concentrations n'entreraient pas dans l'analyse de l'effet sublétal. Si la croissance était modifiée seulement près des concentrations finalement mortelles, les observations de l'effet sublétal correspondant à la partie supérieure de la courbe dose-effet seraient manquantes ou rares, et l'estimation du paramètre de toxicité sublétale pourrait être inappropriée ou mal fondée. Une telle situation serait relativement rare, mais elle pourrait se produire. On se prémunirait contre en intégrant dans le plan d'expérience un plus grand nombre de concentrations, plus rapprochées les unes des autres. Environnement Canada conseille de 8 à 10 concentrations dans les essais de mesure d'un double effet.

L'option 2, mauvais usage parfois usité dans le passé, comporte le problème déjà mentionné de traitement déséquilibré des organismes morts dans les répétitions où la mortalité a été totale, par rapport aux répétitions où la mortalité a été partielle. Du moins, cela signifierait des effectifs déséquilibrés dans les répétitions, tandis que les méthodes d'analyse pourraient avoir été conçues pour des effectifs équilibrés. À part cela, toute analyse semblerait désespérément compromise par le fait qu'il existe deux catégories de données.

L'option 3 pourrait comporter le problème mathématique commun d'effectifs déséquilibrés dans les répétitions et (ou) les concentrations, problème qui peut être résolu grâce aux méthodes statistiques appropriées.

Au niveau de la recherche, Wang et Smith (2000) ont proposé une approche potentiellement supérieure. Elle diffère des options précédentes, mais elle est statistiquement complexe et sa mise au point n'est pas terminée. La modélisation tient compte *à la fois* de la mortalité et des effets sublétaux et elle permet d'estimer une CI p fondée sur les deux effets, ainsi que ses limites de confiance. Les auteurs admettent que l'ajustement de leur modèle n'était pas tout à fait satisfaisant. Ils mentionnent que des modèles « plus complexes » pourraient convenir davantage ; apparemment, leur méthode statistique déjà complexe n'est pas une solution immédiate aux difficultés mentionnées dans la présente section.

8.3 Le nombre de descendants en tant qu'effet sublétal

Repères

- *Dans un essai mesurant un double effet (mortalité et nombre de descendants), l'évaluation de l'effet combiné dans une approche fondé sur la biomasse est un choix qui s'offre à l'analyse.*
- *L'autre méthode légitime, une analyse séparée de l'effet sublétal sur la reproduction (par ex. chez* Ceriodaphnia*) est plus complexe que les essais mesurant la croissance. C'est que le nombre de descendants dépend, en partie, de la durée de survie des parents.*
- *Une approche appropriée peut se fonder sur une disposition en tableaux chronologiques du nombre moyen de nouveaux descendants par*

parent vivant durant la période visée d'inspection. La méthode mérite d'être normalisée, grâce à un progiciel commode.

Si au moyen d'un essai à double objectif, on mesure le nombre de descendants (la « reproduction ») comme effet sublétal, une autre complexité s'ajoute à celles que nous avons décrites dans le § 8.2. C'est la situation illustrée par l'essai sur la reproduction de la puce d'eau, *Ceriodaphnia* (EC, 1992a), mais également celle qui s'applique à la reproduction des vers de terre et des collemboles (EC, 2004a,b et 2007). Dans l'essai avec *Ceriodaphnia,* chaque daphnie adulte entreprend l'essai dans un récipient séparé et, en conséquence, elle représente une répétition à une concentration donnée. Le nombre de jeunes qu'elle aura engendrés à la fin de l'essai est la donnée relative à l'effet sublétal qu'utilise l'analyse statistique de la répétition. (En outre, la mortalité des daphnies adultes est analysée par des méthodes quantiques pour estimer un paramètre de toxicité tel que la CL 50 ou la CL 25.)

La méthode d'essai d'Environnement Canada fonde l'analyse et l'interprétation sur ce dénombrement direct du nombre réel de jeunes engendrés dans chaque répétition, que le parent ait survécu ou non, ce qui est approprié au concept de biomasse.

8.3.1 Interrelation entre la mortalité et la reproduction

Si une daphnie meurt avant de s'être reproduite, le nombre de jeunes dans cette répétition est nul. Cependant, si la daphnie vit assez longtemps pour se reproduire, le nombre observé de jeunes dépend en partie de la longévité du parent, puisque, normalement, celui-ci engendrerait des générations répétées. Ainsi, la mesure apparemment nette d'un effet sublétal dans l'essai avec *Ceriodaphnia* (nombre de jeunes engendrés dans un récipient pendant l'exposition) a, de fait, intégré en elle celle de la mortalité du parent.

Ce type particulier d'intégration avec la mortalité n'est pas un facteur dans le paramètre de toxicité sublétale relatif au poids des larves de tête-de-boule (§ 8.2). Dans des essais avec ces poissons, la mortalité a déterminé le nombre de larves présentes à la fin de l'essai. Cependant, pour l'option 1 recommandée, le critère relatif à un point de donnée sur la toxicité sublétale était indépendant — si une larve vivait jusqu'à la fin de l'essai, son poids contribuerait aux données observées sur la toxicité sublétale, mais, si elle mourait, sa contribution aux données sur le poids serait nulle. Le taux de mortalité dans un groupe n'a pas influé sur l'amplitude du point de donnée (le poids moyen), à l'exception possible de l'« interaction entre les maigrichons et les dodus ».

Lorsque le nombre de jeunes est l'effet mesuré, il existe une interaction avec la mortalité, contrairement à la situation où l'on pèse les têtes-de-boule. La mortalité des parents influe sur le nombre de jeunes dans l'ensemble de données, c'est-à-dire que la quantification des observations d'un effet sublétal est façonnée par le taux de mortalité. Compte tenu de cela, on peut envisager, pour l'essai de reproduction chez *Ceriodaphnia,* les trois options pour l'analyse des données (§ 8.2.1). Les analyses statistiques actuelles de la reproduction des daphnies reposent sur l'hypothèse d'une distribution normale des données, mais elles devraient se fonder sur une distribution de Poisson.

Conceptuellement et mathématiquement, l'option 3 convient à l'analyse des résultats de l'essai avec *Ceriodaphnia,* si l'approche fondée sur la biomasse est reconnue comme un critère approprié d'effet. L'analyse tient compte d'effectifs nuls, faibles et nombreux chez la progéniture, sans égard à la durée de survie des parents. Cette méthode est, en effet, pratique courante dans les essais de reproduction avec *Ceriodaphnia* effectués au Canada et aux États-Unis

Nous avons décrit l'option 2 dans le § 8.2 et nous n'en tiendrons pas compte ici pour les raisons précédemment exposées.

L'option 1, que nous ne recommandons pas, buterait sur la difficulté supplémentaire décrite précédemment, selon laquelle la mortalité des parents ne peut pas être facilement séparée de l'effet sublétal, bien que cette séparation soit la conséquence directe de cette option. Si un adulte meurt avant d'engendrer, l'effectif nul de sa progéniture noté pour cette répétition ne représente pas un effet sublétal sur la reproduction, mais, plutôt, la mortalité. De même, si un parent meurt prématurément, le faible nombre de descendants refléterait cette mortalité, plutôt qu'une fatigue des

mécanismes de reproduction[60]. Une nouvelle approche potentielle à ce problème est exposée dans le § 8.3.2.

8.3.2 Analyse séparée de la reproduction

Hamilton (1986) a étudié de façon pénétrante le problème de l'estimation de la progéniture dans les essais de toxicité avec *Ceriodaphnia.* Il a rédigé une méthode potentiellement favorable à l'option 1. Cette approche, qui consiste à séparer l'effet sublétal de la mortalité, mérite qu'on l'évalue en vue d'un emploi futur. Il est surprenant que cela n'ait pas déjà été fait.

Hamilton (1986) a utilisé les résultats d'un essai réel avec *Ceriodaphnia* pour prouver les erreurs systématiques (biais) qui se manifestent si on fonde le nombre de jeunes soit sur le nombre initial d'adultes, soit sur le nombre d'adultes survivants à la fin de l'essai. Une solution serait de présenter dans un tableau le nombre de jeunes engendrés par chaque adulte vivant, à chaque inspection (l'essai dure habituellement sept jours, pendant lesquels on compte et retire journellement la progéniture des récipients). On calcule la moyenne journalière par adulte dans toutes les répétitions à une concentration donnée.

Cette approche n'est valide que s'il n'existe pas de corrélation entre la mortalité et la reproduction. Si la mortalité prochaine ralentissait la reproduction, l'interaction pourrait disqualifier cette méthode. Hamilton (1986) a prouvé que cette corrélation était négligeable ou absente. Les *Ceriodaphnia* continuaient de se reproduire à un rythme normal jusqu'à leur mort, dans la mesure où cela pouvait être décelé par des moyens statistiques et par des comparaisons graphiques convaincantes. Dans le même temps, l'examen des données par Hamilton a montré que l'approche fondée sur la biomasse reflétait d'abord les mortalités et non les taux de reproduction.

À la fin de l'essai, on a réuni les moyennes journalières (du nombre de jeunes par parent) pour obtenir le nombre moyen total de jeunes par adulte à chaque concentration. Ces données de base représentaient une estimation relativement non biaisée de la performance de reproduction. Hamilton (1986) a prouvé, à partir des données ajustées, que la méthode permettait de déceler des modifications des performances de reproduction, à part toute influence de la mortalité. Pour mesurer la variation à chaque concentration, Hamilton a recommandé des techniques *bootstrap.*

L'approche exposée par Hamilton (1986) semblait bien étayée et bien justifiée. Nous la recommandons comme méthode à mettre au point pour les essais de mesure d'un double effet où la reproduction est l'effet sublétal, comme dans l'essai avec *Ceriodaphnia.* Cette approche, fondée sur l'option 1, séparerait l'effectif de la descendance comme un effet individuel et elle est semblable à celle que l'on utilise en épidémiologie humaine pour étudier le temps prévu de survie à partir d'une cause donnée (disons une crise cardiaque), abstraction faite des effets de causes concurrentes de mortalité. Elle est également semblable aux méthodes utilisées en biologie des pêches pour faire abstraction de l'effet de la mortalité par pêche, de sorte que l'on peut décrire les caractéristiques naturelles des populations halieutiques (Ricker, 1958).

Il faudrait préciser et normaliser le mode opératoire de cette méthode d'analyse, puis développer un progiciel commode. La méthode pourrait s'appliquer à tout essai de mesure d'un double effet dans lequel une réponse cumulative a été utilisée pour chaque animal et dans lequel la mortalité pourrait survenir prématurément. Le paramètre préféré de toxicité serait la CI p. Hamilton (1986) a recommandé, outre l'analyse mathématique, le tracé des graphiques du nombre de jeunes engendrés journellement dans chaque répétition, pour évaluer la séparation de la mortalité et de la reproduction.

---

60. La position correspondant à l'option 1 serait encore plus intenable si on tentait naïvement d'exprimer les résultats correspondant à une concentration particulière comme le nombre moyen de jeunes engendrés par adulte. Faire ce calcul, c'est toujours s'attirer des ennuis. Il serait difficile, si un ou plusieurs adultes mouraient pendant l'essai, de calculer une moyenne réaliste sans tenir compte de la durée de survie. Si on divisait le nombre total de jeunes par le nombre initial de parents, la moyenne serait systématiquement ramenée vers le bas (par ex. un adulte mort sans descendance au premier jour d'exposition figurerait toujours dans le calcul comme s'il était un parent fécond). Si on divisait le nombre de jeunes par le nombre de parents survivants à la fin de l'essai, la moyenne serait poussée vers le haut, au-dessus d'une valeur réaliste. (Par exemple, un adulte mort une heure avant le dernier dénombrement aurait probablement engendré tous les jeunes qu'il était capable d'engendrer, mais ne compterait pas dans le calcul du nombre moyen de jeunes par adulte. Comble de l'absurdité, si tous les adultes avaient engendrés leur descendance, mais étaient morts une heure avant le dénombrement final, un grand nombre de jeunes serait attribué à un nombre nul d'adultes.)

Cette option 1 ou méthode « de séparation » pourraient être une solution de rechange à l'analyse fondée sur la biomasse, utilisée dans l'essai d'Environnement Canada avec *Ceriodaphnia*. Cependant, les deux approches sont identiques si tous les adultes survivent jusqu'à la fin de l'essai.

8.4 Résumé et recommandations

Il va de soi que les essais conformes aux méthodes publiées par Environnement Canada doivent utiliser la méthode prescrite d'analyse. Dans d'autres situations, le choix d'une méthode appropriée et d'une analyse statistique doit être faite par l'expérimentateur pour répondre aux besoins de l'étude.

Pour l'analyse et l'interprétation des résultats des essais de mesure d'un double effet, il existe deux grandes options légitimes. La première (l'option 1) consiste à séparer l'effet sublétal de l'autre effet (d'habitude la mortalité) et de l'analyser séparément. Cette séparation des effets pourrait être plus instructive, techniquement.

La seconde approche consiste à combiner les deux effets dans un type d'analyse fondée sur la biomasse. Elle pourrait augmenter l'effet toxique apparent, et les résultats pourraient mieux prévoir les effets écologiques globaux du monde réel. L'approche fondée sur la biomasse pourrait être appropriée pour des essais ou des objectifs particuliers. L'application générale de cette option ne trouve cependant pas beaucoup d'appui parmi les expérimentateurs canadiens en exercice (Schroeder et Scroggins, 2001).

On devrait éviter les méthodes qui combinent partiellement deux effets.

Dans les essais de toxicité mesurant un double effet, notamment la taille atteinte par les organismes, l'option 1 semble préférable. Elle utilise la taille moyenne des survivants et elle permet des analyses statistiques « propres ». De telles observations se prêtent à la plupart des méthodes statistiques communes qui tiennent compte d'effectifs inégaux dans les répétitions. On ne devrait pas ignorer la mortalité ni d'autres effets, mais les signaler convenablement après analyse par des méthodes quantiques.

Dans les essais de mesure d'un double effet (notamment le nombre de descendants), l'approche fondée sur la biomasse est une solution appropriée de rechange pour l'analyse et l'interprétation. Il serait souhaitable qu'Environnement Canada élabore et normalise une telle approche, qui isolerait l'effet sublétal (la reproduction) par le nombre moyen de jeunes par parent, présenté dans des tableaux créés pour chaque période d'inspection et dont le total aura été calculé pour l'ensemble de l'essai (v. la description donnée dans le § 8.3.2).

Quelques concepts et outils de statistique

La plupart des expérimentateurs auront déjà été initiés à la statistique, et le présent guide ne prétend pas jouer ce rôle d'initiation. Cependant, nous avons défini dans le glossaire, pour des motifs de commodité, des termes de statistique se rapportant à la toxicologie. En outre, des rudiments de mathématiques, en rapport avec les analyses de toxicité, sont exposés au début de la présente section ; vers la fin de cette dernière, il sera question de certaines méthodes mathématiques souvent utilisées.

9.1 Distributions normales et binomiales

Repères

- *Les distributions normales et binomiales sont des caractéristiques fondamentales des essais de toxicité quantitatifs et quantiques, respectivement. Quand on a affaire à de grands nombres et à des proportions s'approchant de 0,5, les courbes binomiales tendent à ressembler aux courbes normales.*

Les distributions normales sont fondamentales pour beaucoup de résultats des essais de toxicité, à l'instar de la plupart des domaines de la biologie. Beaucoup de tests statistiques reposent sur l'hypothèse de la normalité des données, particulièrement les résultats des essais quantitatifs de toxicité sublétale (section 6). De même, la distribution binomiale est fondamentale pour les données quantiques (section 4). Pour un grand nombre d'observations et pour les proportions proches de 0,5, la distribution tend à ressembler à la distribution normale.

9.1.1 Courbes normales

Nous décrivons, dans le glossaire, les caractéristiques de la distribution normale et nous en donnons une représentation dans la fig. 20, en prenant les statures comme exemple.

La distribution normale caractéristique du graphique supérieur de la fig. 20 montre que la plupart des statures se regroupent autour de la moyenne. Plus on s'éloigne de cette dernière, moins les observations sont nombreuses. L'histogramme peut être représenté par la courbe normale classique en forme de cloche. Les courbes en forme de cloche ne sont pas toutes normales ; pour être qualifiée de normale, la distribution doit satisfaire à une formule assez complexe (Zar, 1999). Des tests usuels permettent d'établir si un ensemble de données remplit les exigences à cette fin (§ 7.3).

Le graphique inférieur de la fig. 20 montre comment la grandeur de l'écart type (σ, sigma) détermine la forme de la courbe normale. X y représente la variable mesurée, dont la moyenne, dans ce cas est de 0. L'axe vertical représente la fréquence f (ou la probabilité) de réalisation. Plus l'écart type est grand, plus la courbe est large et aplatie. La variation de la valeur de la moyenne déplacerait la courbe vers la gauche ou la droite, mais n'en modifierait pas la forme. Les courbes normales sont toujours symétriques, bien qu'une distribution asymétrique puisse résulter de la superposition d'une distribution normale sur une autre distribution (c'est-à-dire de la combinaison de deux ensembles de données dont les moyennes sont différentes).

9.1.2 Distributions binomiales

Les distributions binomiales sont très importantes en écotoxicologie parce qu'une grande partie des données est du type « tout ou rien ». Beaucoup d'essais consistent à dénombrer les organismes morts à la fin de l'expérience par rapport au nombre total exposé. On peut qualifier de telles données de *binomiales,* de *binaires* ou de *quantiques* (v. le glossaire). Dans la fig. 21, nous montrons des histogrammes de données binaires.

Le graphique de gauche de la fig. 21 montre une distribution symétrique quand la probabilité est de 0,5. Si on réduit la probabilité de l'événement, la distribution devient asymétrique, comme dans les

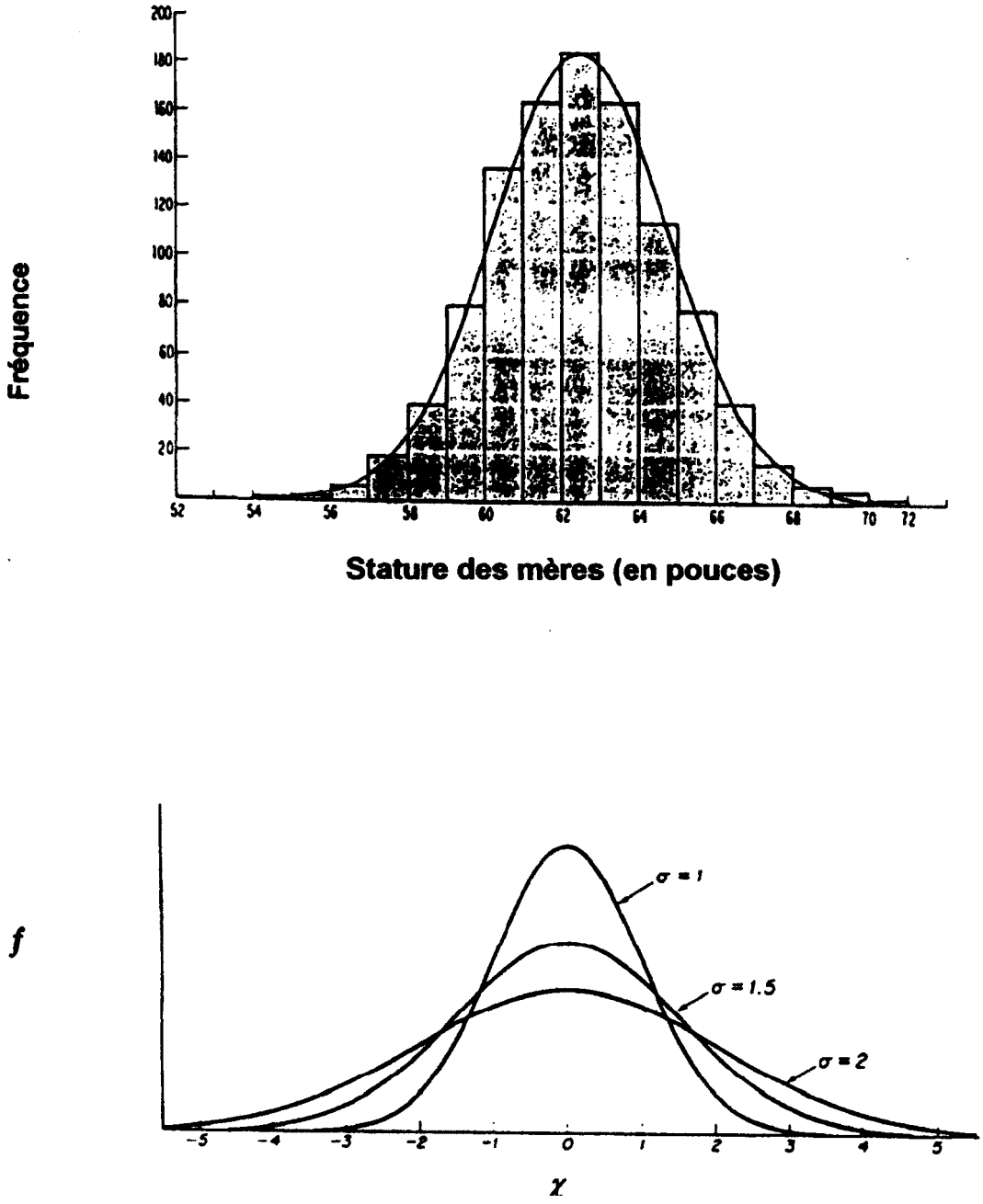


Figure 20. — Distributions normales. Le graphique du haut montre la répartition des statures de 1 052 personnes, ajustées à une courbe normale en forme de cloche. Dans le graphique du bas, χ représente la variable mesurée, avec une moyenne de 0. L'axe vertical représente la fréquence (f). La forme de la courbe est commandée par l'écart type (σ). D'après Snedecor et Cochran (1980) et Zar (1974).

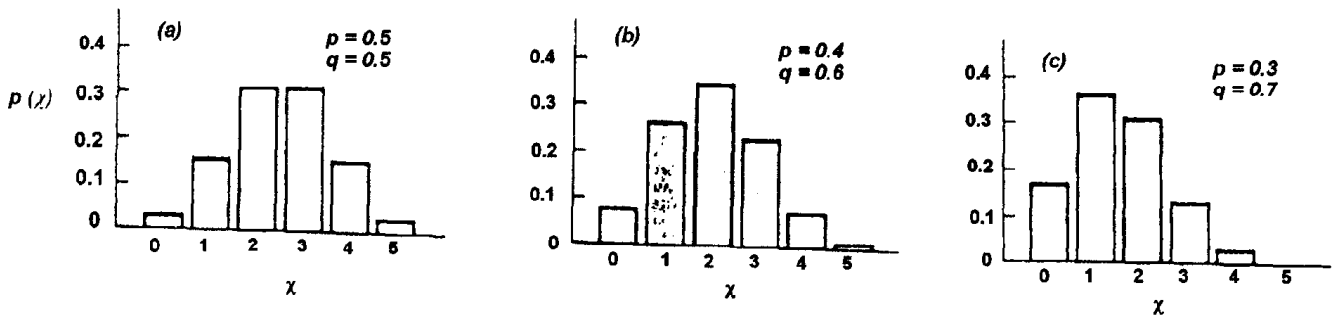


Figure 21. — Distributions binomiales. Voici les distributions consécutives à cinq essais d'un événement binomial (par ex. « mortalité » ou « absence de mortalité » chez cinq puces d'eau dans une enceinte). La probabilité de survenue de l'événement (« mort ») est de *p,* tandis que la probabilité de non-survenue (« vivant ») est de *q*. L'axe horizontal *(x)* sous chaque histogramme représente 0, 1, 2, etc. réalisations de la première probabilité (c'est-à-dire aucune mortalité, un mort, etc. chez les cinq organismes en 5 essais ou répétitions). L'axe vertical représente la fréquence de ces manifestations. Dans le graphique de gauche (*a*), les probabilités de la réalisation ou non de l'événement sont égales, et la distribution est symétrique. Dans le graphique *b* et *c*, les probabilités de survenue de l'événement sont réduites(la mort est moins probable), et les distributions sont asymétriques. D'après Snedecor et Cochran (1980).

graphiques *b*) et *c*). La fréquence correspondant aux barres du côté gauche des histogrammes augmente, notamment pour les réalisations nulles sur 5 essais (X = 0, soit aucune mortalité). Par conséquent, la fréquence diminue (barres de droite des histogrammes), notamment la disparition de cinq réalisations sur cinq essais (X = 5 ou la mort des cinq organismes).

À la lecture de la fig. 21, il est facile de voir qu'avec des échantillons plus gros (disons d'au moins 25 individus) et avec $p \approx 0{,}5$, la distribution binomiale des essais (ou des organismes) prendrait la forme générale d'une distribution normale (fig. 21). Beaucoup d'observations se regrouperaient près de la proportion de 0,5, et elles seraient de moins en moins nombreuses à mesure que les proportions s'éloigneraient de cette valeur en tendant vers 0 ou 1,0. Si p s'écartait sensiblement de 0,5, la distribution normale serait une mauvaise approximation de la binomiale. Selon la valeur de *p,* des centaines d'observations binaires pourraient devoir être nécessaires pour obtenir une distribution semblable à la normale. Cette caractéristique est en rapport avec les hypothèses de normalité utilisées pour estimer la fécondation dans les essais de toxicité employant des oursins et des salmonidés (EC, 1992f ; 1998a).

9.2 Échantillons et populations

Repères

- *Les essais de toxicité utilisent toujours un échantillon d'organismes, et une sélection aléatoire est essentielle si l'on veut que l'échantillon soit représentatif de la population se trouvant dans un réservoir d'attente.*
- *Il est rare que l'on tente de déterminer si un essai particulier de toxicité est représentatif des populations beaucoup plus nombreuses d'organismes vivant librement (sauvages). Cependant, la plupart des essais délibérés de validation sur le terrain confirment que les concentrations toxiques déterminées en laboratoire sont de bons prédicteurs des effets nocifs pour les communautés naturelles (sauvages).*

Les expérimentateurs effectuent des essais de toxicité sur un *échantillon* d'organismes. Ils pourraient prélever un échantillon dans un réservoir contenant beaucoup d'organismes. Tous les organismes du réservoir

pourraient être considérés comme une population. L'expérimentateur suppose que l'échantillon est typique des organismes en attente dans le réservoir ; c'est pourquoi un processus de sélection aléatoire des échantillons est important.

Le paramètre de toxicité estimé à la faveur de l'essai et ses descriptions statistiques caractérisent toujours l'*échantillon.* Les tests et les descriptions statistiques tiendront compte de la taille de l'échantillon et de la variation des observations, dans le cadre du processus d'estimation. Un grand échantillon est susceptible de produire un paramètre de toxicité plus précis. D'où la concurrence qui, dans le plan d'expérience, oppose le désir de travailler sur un grand échantillon pour obtenir plus de précision, d'une part, et le désir de travailler avec de petits échantillons pour réduire la taille de l'appareillage, la quantité de substrat et le temps nécessaire à la vérification des effets.

Il est habituellement raisonnable de poser comme hypothèse que le paramètre de toxicité pour l'échantillon représente aussi la population. Cependant, si l'expérimentateur s'y est mal pris pour l'échantillon (disons qu'il n'a prélevé que de gros organismes), toute constatation statistique s'appliquerait à l'échantillon, mais non à la population dans le réservoir d'attente.

À un échelon plus général, se trouve une hypothèse implicite, dont il ne sera pas question ici, selon laquelle les organismes se trouvant dans le réservoir d'attente et le paramètre de toxicité de l'échantillon représentent une population beaucoup plus grande telle que la totalité des organismes sauvages de l'espèce employée. Une telle hypothèse est rarement testée pour un ensemble d'essais de toxicité en laboratoire, et cela doit être reconnu par les utilisateurs des données expérimentales. Il est donc essentiel de présenter des renseignements sur l'échantillon d'organismes testés, tels que le contexte génétique, l'historique de l'élevage et la taille. Ces éléments d'information sont exigés dans les méthodes publiées par Environnement Canada.

Cependant, il existe une masse importante de renseignements sur la validation sur le terrain des essais de toxicité en laboratoire. Le travail de terrain dans les parages de certaines usines canadiennes de pâte à papier a montré que les effets observés dans la nature correspondaient aux prévisions découlant des essais effectués en laboratoire (Scroggins *et al.,* 2002) et que les évaluations en laboratoire étaient également utiles pour prévoir les effets des mines de métaux (Sprague, 1997). Il y a eu un nombre appréciable de programmes de recherche sur le terrain pour associer les effets subis dans les communautés aquatiques naturelles (sauvages) et les résultats d'essais en laboratoire ainsi que d'expériences similaires employant des communautés contrôlées (mésocosmes). Ayant assuré un examen majeur de cette recherche, Environnement Canada a conclu que, dans la plupart des cas, les essais en laboratoire étaient de bons prédicteurs des effets dans les habitats naturels (EC, 1999a).

Des détails encore plus pertinents peuvent être trouvés dans le glossaire sous les vedettes *échantillon, population, unité d'échantillonnage, unité expérimentale, traitement, répétition, échantillonnage au hasard, erreur d'échantillonnage* et *précision.*

9.3 Signification statistique par opposition à signification biologique

Repères

- *Les tentatives de définition des degrés d'effet toxique biologiquement significatif sont relativement peu nombreuses. Idéalement, dans un test d'hypothèse, un tel niveau devrait être défini préalablement à l'essai de toxicité. L'essai et son analyse statistique pourraient ensuite être convenablement planifiés, de façon à évaluer la signification biologique. Le résultat de l'essai serait que l'on a observé ou non un effet biologique néfaste avec une certitude de 95 %.*
- *Actuellement, la signification statistique des effets remplace généralement la signification biologique, par défaut, mais les deux notions ne correspondent pas nécessairement sans un plan approprié d'expérience.*

La signification statistique des résultats est un thème récurrent du présent document et de l'écotoxicologie. Cela est particulièrement vrai dans le test d'hypothèse(s). Presque universellement, on choisit comme critère la probabilité de 5 % en vertu de laquelle toute différence sera attribuable au hasard. Si une

différence observée est suffisamment importante pour n'être due au hasard qu'une seule fois sur 20 (ou moins), on la considère comme significative. Ce niveau de signification signifie que si 20 essais de toxicité étaient effectués sur une substance inoffensive, on devrait s'attendre à ce que les résultats d'un essai présentent une différence significative par rapport au témoin (erreur α ou de première espèce conduisant à conclure, à tort, à l'existence d'une différence ; § 7.2.2).

La lacune de l'approche générale est la suivante : la signification biologique est rarement définie, de sorte que ce concept ne peut pas être intégré dans le plan d'expérience de l'essai de toxicité. Quand l'essai est terminé, la signification biologique et la signification statistique n'ont pas besoin d'avoir de relation particulière l'une avec l'autre. Dans le § 7.1.2, nous avons mentionné que la concentration sans effet statistique était associée à des effets biologiques sublétaux qui, en moyenne, étaient plus graves de 14 % que chez le témoin et qui pouvaient même être plus graves de 38 % (Crane et Newman, 2000).

La bonne façon de faire serait que le biologiste ou le toxicologue décide *dès le début du processus* de ce qui constitue un effet écologiquement significatif dans contexte particulier (survie, croissance, taux de reproduction, etc.) et qu'il en informe le statisticien. À son tour, ce dernier intégrerait ce degré d'effet dans le test d'hypothèses et informerait le toxicologue des conditions à remplir concernant le nombre d'échantillons, de répétitions, d'organismes plus un certain degré de variation. Après analyse des résultats de l'essai, on conclurait qu'un effet écologiquement significatif a été (ou n'a pas été) démontré avec une certitude de 95 %. (Cela suppose que l'on a fixé la valeur de β à 0,05.)

Une décision rare et louable sur la signification biologique a été prise au Canada, dans le cadre du Programme d'immersion en mer. Les critères d'une différence biologique significative ont été fixés à 20-30 % de différence par rapport au témoin dans certains essais sur un sédiment (Porebski et Osborne, 1998 ; Zajdlik *et al.*, 2000 ; v. le § 7.2.5). [Les constatations doivent aussi être statistiquement significatives, bien sûr.]

Cette question de jugement est mise en valeur dans une autre approche importante de l'écotoxicologie, l'estimation ponctuelle (section 6). Le paramètre de toxicité qu'est la CI p peut prendre toute valeur de *p* choisie par l'expérimentateur. La CI 25, qui correspond à une réduction de 25 % des performances (par rapport au témoin), est parvenue à être généralement reconnue comme paramètre de toxicité dont la signification écologique est acceptable (§ 6.2.4).

La décision sur ce qui constitue un effet écologique significatif doit se fonder sur des critères biologiques et le jugement de l'expérimentateur. Le degré choisi d'effet pourrait varier selon le type d'effet. Peut-être une diminution de 50 % du nombre d'œufs ne serait-elle pas considérée comme de la plus haute importance écologique, mais un ralentissement de 10 % de la croissance des individus pourrait être considéré comme très important.

Faute de décisions initiales sur la signification biologique, l'écart potentiel entre cette dernière et la signification statistique peut aller dans un sens comme dans l'autre. Un effet statistiquement significatif pourrait être minime et ne causer aucune inquiétude sur le plan biologique. Cependant, un effet qui n'est pas statistiquement significatif pourrait être biologiquement grand, être très inquiétant, se manifester dans un essai présentant une grande variabilité interne. Cette contradiction est peut-être d'une plus grande importance pratique. L'expérimentateur est déchiré entre les résultats statistiques et la responsabilité de signaler un effet biologique majeur. Ce qu'il faut éviter, c'est de tomber dans la phraséologie des premiers jets de thèses de maîtrise : « bien que statistiquement non significatif, l'important changement de... montre que... » De fait, dans un cas comme celui-là, l'expérimentateur n'a pas montré qu'il y avait de changement, *quel qu'il soit,* par rapport au témoin.

Paine (2002) décrit de façon excellente les conflits généraux relatifs aux programmes de surveillance des effets sur l'environnement, dans le cadre de l'approche actuelle aux plans d'expérience :

> « Les effets significatifs pour l'environnement peuvent ne pas être statistiquement significatifs, et les effets statistiquement significatifs peuvent ne pas être significatifs pour l'environnement. L'ampleur des effets significatifs pour

> l'environnement est difficile à définir, parce qu'elle dépend d'enjeux et de valeurs environnementaux, sociologiques, politiques et économiques. Par conséquent, nous traitons souvent comme équivalentes, implicitement ou par nos actions, la signification environnementale et la signification statistique. Les discussions et les décisions juridiques, réglementaires et gestionnelles se fondent souvent sur la signification statistique de résultats ou d'effets. De façon plus générale, les articles de journaux ainsi que les rapports de consultants ou de fonctionnaires ne communiquent que la signification statistique des effets (par ex. "la fécondité des poissons a été significativement plus faible dans la région touchée que dans la région témoin")... » [Traduction]

Paine (2002) formule trois recommandations pour que nous ne nous sentions pas parfois écartelés entre la signification statistique et la signification biologique. Certes, les deux premières devraient être suivies par les expérimentateurs qui font rapport de leurs travaux. La deuxième est l'essence de l'argument que nous venons de présenter.

(1) Signaler l'ampleur des effets et ses limites de confiance et ne pas simplement affirmer que les effets étaient statistiquement significatifs.

(2) S'efforcer de définir les effets écologiquement significatifs, quelque difficile que cela puisse être.

La troisième recommandation de Paine, cesser d'être obsédé par la signification statistique, serait le mieux satisfaite grâce à la planification des essais de toxicité pour que les résultats statistiques aient une signification directe pour l'effet biologique.

9.4 Régression inverse

L'expérimentateur devrait être conscient que, dans l'essai usuel d'écotoxicité, l'estimation du paramètre de toxicité et de ses limites de confiance pose un problème statistique complexe. Cette complexité relève du traitement statistique, de sorte que l'expérimentateur n'a à prendre aucune action correctrice. Cependant, cette complexité explique pourquoi il faut utiliser des méthodes statistiques spécifiques et pourquoi les limites de confiance sont souvent asymétriques.

Repères

- *Dans un essai de toxicité, la concentration est d'abord la variable indépendante. L'essai mesure la variation d'après l'effet biologique, la variable dépendante. L'estimation du paramètre de toxicité et des limites de confiances est, cependant, reconvertie en concentration. Cela entraîne des complexités statistiques.*
- *L'expérimentateur est en grande partie ignorant de la complexité de l'analyse statistique, mais cela explique pourquoi il faut utiliser des programmes particuliers d'analyse des données sur la toxicité et pourquoi les limites de confiance peuvent être asymétriques.*
- *L'inversion ne s'applique pas aux paramètres de toxicité exprimés en temps tels que le* temps efficace 50 *(TE 50), puisque les observations et les calculs se fondent sur une variation à l'échelle du temps.*

Habituellement, l'expérimentateur fixe les concentrations quand il organise l'essai de toxicité, faisant de la concentration la variable indépendante. Le degré d'effet biologique constaté chez les organismes est mesuré en tant que variable dépendante. Cela instaure un conflit fondamental entre le plan d'expérience et les paramètres de toxicité recherchés. En un mot, les concentrations finissent par être traitées comme si elles étaient la variable dépendante. La détermination du paramètre de toxicité est inversée, pour estimer la concentration nécessaire pour causer un degré d'effet biologique fixé par l'expérimentateur, c'est-à-dire des paramètres tels que la CE 50, la CI 25 et leurs limites de confiance. L'inversion entraîne des complexités statistiques dans les programmes d'analyse des données.

Les concentrations d'essai fixées a priori sont censées être invariables. Les observations expérimentales du degré d'effet sont exposées à une variation expérimentale de l'effet véritable. Si on calcule une relation linéaire entre les deux, la variabilité de cette relation continue d'être en fonction de l'effet biologique mesuré. Cette relation linéaire, avec sa variation sur l'axe des effets sert à prévoir les paramètres de toxicité et leurs limites de confiance sur l'autre axe, c'est-à-dire

l'axe des concentrations. Par exemple, une CE 50 (*concentration*) et les concentrations marquant ses limites de confiance seraient estimées à partir de la droite ajustée et de la variation des *effets* que cause cette concentration (v. la fig. 7).

Le conflit est moins évident dans le test d'hypothèse(s). L'estimation d'un paramètre de toxicité est simple, parce qu'elle se fonde sur la variation observée de l'effet pour déterminer le traitement causant un effet significativement différent de l'effet observé chez le témoin. Cependant, pour les limites de confiance, c'est une estimation inversée qui prend la relève : exprimées en unités de concentrations, elles sont calculées à partir de la variation de l'effet.

La commutation réciproque des variables dépendantes et indépendantes peut être décrite comme une *estimation inverse des paramètres de toxicité* et des limites de confiance. Comme nous l'avons mentionné, l'inversion est intégrée dans les programmes statistiques, de sorte que l'expérimentateur n'en a pas conscience. Elle reste néanmoins une complexité dans les opérations statistiques de la plupart des essais employés en écotoxicologie.

Un effet commun de l'estimation inverse est révélé par l'exemple d'un essai de toxicité quantique hypothétique dans la fig. 7. À toute concentration donnée, les limites de confiance sont symétriques, verticalement, parce qu'elles sont calculées en *effets observés* aux concentrations fixées. Cependant, les limites de la CE 50, parallèles à l'axe horizontal des concentrations, sont habituellement asymétriques en raison de l'inversion des calculs. L'asymétrie est manifeste si l'on pose une règle horizontalement sur la fig. 7, à la hauteur de l'effet de 50 % ou à tout autre taux d'effet. L'asymétrie est particulièrement évidente près des extrémités de la droite des probits, où l'une des limites ou les deux peuvent parfois s'en éloigner beaucoup, presque à l'infini même.

Les estimations inversées s'appliquent quand toute technique de régression, qu'elle soit linéaire ou non, est appliquée aux essais habituels de toxicité. Les progiciels ordinaires de statistique (ceux qui ne s'appliquent pas à la toxicologie) n'offrent pas l'option « standard » permettant de traiter ce phénomène, par ex. dans l'estimation de limites de confiance. Voilà l'une des raisons pourquoi il faut utiliser un programme spécialement conçu de régression probit, plutôt qu'une simple méthode d'ajustement de la droite fondée sur les moindres carrés, pour estimer la CE 50. Bien que Nyholm *et al.* (1992) aient fourni une formule pour estimer les limites de confiance d'un paramètre de toxicité estimé par régression linéaire ordinaire, elle ne semble pas encore avoir été intégrée dans les progiciels nord-américains d'écotoxicologie. On peut trouver des formules servant au même usage général dans Draper et Smith (1981) et dans d'autres manuels sur la régression.

La régression inverse ne s'applique pas aux essais quantiques d'estimation du TE 50 ou du TL 50 (v. la section 5). Dans ce cas, on estime le paramètre de toxicité et ses limites de confiance en unités de la variable dépendante, le temps. L'approche directe est statistiquement propre, s'ajoutant aux autres avantages de l'emploi du TE 50 comme paramètre de toxicité.

L'autre approche générale au problème de l'estimation inverse consiste à reparamétrer l'équation reliant l'effet à la concentration (v. le § 6.5.12). Environnement Canada l'a fait dans de récentes méthodes d'essai de toxicité d'un sol (EC, 2004a, b et 2007 ; § 6.5.7 et 6.5.8).

9.5 Différences significatives entre les CE 50

Repères

- *On peut évaluer des différences significatives entre deux paramètres quantiques de toxicité (des CE 50) à partir de leurs limites de confiance. La comparaison simple est analogue à l'erreur type de la différence.*
- *Il semble faisable d'employer une méthode mathématique supérieure pour deux CE 50.*
- *Pour tester les différences entre plusieurs CE 50, on pourrait utiliser l'analyse de variance classique pour la situation inhabituelle dans laquelle on possède des répétitions.*
- *Il semble possible de mettre au point une formule mathématique exclusive pour déterminer si une différence significative a existé entre plusieurs CE 50, mais, à l'instar d'une analyse de variance,*

elle ne permettrait pas de déterminer la ou les CE 50 qui diffèrent de la sorte.

On peut calculer des différences significatives entre des paramètres de toxicité sans recourir à des méthodes à cette fin, lorsque l'on dispose de données brutes. Cependant ces méthodes n'entrent pas dans le cadre du présent document. Dans le présent paragraphe, nous décrivons des méthodes que l'on peut utiliser à cette fin particulière lorsque l'on ne dispose pas de données brutes.

9.5.1 Paires de CE 50

On peut utiliser des méthodes à cette fin pour comparer la différence statistiquement significative entre deux paramètres quantiques de toxicité.

Aucune superposition des intervalles de confiance. — L'examen des limites de confiance est commode pour déterminer, pour les résultats d'essais quantiques, des différences significatives entre certaines paires de CE 50. Si les intervalles de confiance de ces concentrations ne se superposent pas, les CE 50 sont différentes et on peut les déclarer telles sans autre test statistique. Cependant, la superposition des intervalles de confiance ne dit rien sur le caractère significatif ou non de la différence entre ces paramètres.

Méthode de Litchfield-Wilcoxon. — Pour distinguer deux CE 50, on peut utiliser cette méthode (Litchfield et Wilcoxon, 1949) , analogue à une technique mathématique reconnue, l'erreur type de la différence entre les moyennes (Zar, 1974, p. 105-106), bien que la plupart des manuels de statistique ou de toxicologie n'en traitent pas explicitement. La méthode de Litchfield et Wilcoxon (1949) est analogue à l'obtention d'une seule *estimation groupée de la variance* à partir des variances de deux distributions (Snedecor et Cochran, 1980). Finney (1971, p. 110-111) montre un exemple analogue pour obtenir une seule variance pour la puissance relative, à partir de la somme de deux variances d'une paire de substances. La méthode a été mise en doute par Hodson et al. (1977), mais elle fait partie des méthodes usuelles de la pharmacologie. L'application de la méthode à l'écotoxicologie est décrite par Sprague et Fogels (1977). La méthode est utilisée depuis quelques décennies et semble valable pour des paires d'essais dont les résultats ont des distributions semblables.

Cette méthode approximative est appliquée comme le montre l'équation 7. On pourrait l'employer avec précaution [61] tant qu'une méthode mathématique supérieure n'aura pas été mise au point ni publiée.

$$f_{1,2} = anti\log\sqrt{(\log f_1)^2 + (\log f_2)^2} \qquad (7)$$

Pour comparer deux CE 50 dont les intervalles de confiance se superposent , on calcule la statistique $f_{1,2}$ conformément à l'équation 7. La différence entre les deux serait significative si le quotient (CE 50 maximale) /(CE 50 minimale) excédait la statistique $f_{1,2}$. La valeur f_1 est simplement le quotient entre la limite de confiance et la CE 50 pour un essai donné et on peut la calculer comme suit : [(la limite supérieure de confiance) / (CE 50)] + [(CE 50) / (la limite inférieure de confiance)], le tout divisé par 2. On calcule de même f_2 pour l'autre CE 50. Pour effectuer ce calcul, on peut se procurer un petit programme informatique auprès d'Environnement Canada à North Vancouver [62].

La principale utilisation de l'équation 7 serait probablement de déterminer si deux CE 50 ne sont pas différentes, ce qui éviterait la surinterprétation de la

61. L'équation 9.1, analogue à celle de l'erreur type ($s_{\bar{x}}$)de la différence, dans laquelle $s_{\bar{x}.\text{diff}}$ égale la racine carrée de la somme de ($s_{\bar{x}}$ élevé au carré pour le premier élément) plus ($s_{\bar{x}}$ élevé au carré pour le second élément). L'emploi de cette méthode en écotoxicologie pourrait parfois la déloger de sa base statistique prévue. En pharmacologie, les méthodes classiques permettaient de teste un médicament de puissance inconnue par rapport à un autre, étalon, de puissance connue. Le test de la différence significative entre les puissances, à la manière de l'équation 9.1, exigeait la même pente pour les relations dose-effet des deux matières. Dans les essais de toxicité, les « pentes » des distributions des effets pourraient ne pas être les mêmes, de sorte que la validité de l'emploi de cette méthode est mise en doute. Il est probable que si f_1 et f_2 sont semblables, c'est-à-dire que si les intervalles de confiance sont d'une largeur semblable sur une échelle logarithmique, par rapport à leurs CE 50, cette méthode pour tester la différence significative serait acceptable. Si la méthode *ad hoc* de Zajdlik devient disponible, ce serait la méthode de prédilection, en l'absence de données brutes.

62. Programme de toxicologie, Environnement Canada, Centre des sciences environnementales du Pacifique, 2645, route Dollarton, North Vancouver, BC, V7H 1V2.

variation dont on n'a pas prouvé la réalité[63].

Attention ! La conclusion selon laquelle il existe une différence significative ne s'appliquerait qu'aux deux paramètres particuliers de toxicité qui auront été comparés et elle pourrait ne pas être vraie si des essais supplémentaires étaient effectués. Par exemple, si l'équation 7 a montré que les CE 50 du cuivre différaient de façon significative pour deux espèces de crustacés, cela ne signifierait pas nécessairement que les espèces avaient une tolérance différente : uniquement que ces deux paramètres particuliers de toxicité étaient différents. En outre, on devrait prendre en considération la signification biologique des différences et des causes possibles. Par exemple, la variation des résultats obtenus par différents laboratoires ou en différents moments pourrait mener à des différences statistiquement significatives, mais la signification biologique de la différence pourrait se trouver dans le domaine de la variation inexpliquée, et on devrait la considérer sous cet angle.

Méthode particulière n° 1de Zajdlik

Une méthode mathématique de comparaison de deux CE 50 peut se fonder sur le test Z à deux échantillons, que la plupart des manuels de statistique expliquent (par ex. Zar, 1974, p. 105-106). Hubert (1992) a proposé la méthode générale à utiliser dans la comparaison de deux CE 50. Cela pourrait être une méthode utile, une fois que ses étapes auront été décrites par Zajdlik (en préparation). L'équation 8 donne la formule de calcul.

$$Z = \frac{\log CE\,50_1 - \log CE\,50_2}{\sqrt{\left[s^2_{\bar{x}(\log CE50_1)} + s^2_{\bar{x}(\log CE50_2)}\right]}} \qquad (8)$$

$s_{\bar{x}}$ représente l'erreur type, c'est-à-dire les erreurs types du premier et du second logarithmes des CE 50.

Au moment d'écrire ces lignes, la méthode de calcul de $s_{\bar{x}}$ (et, de là, son carré) reste à préciser.

Si | Z | > 1,96, alors les deux CE 50 diffèrent significativement au niveau de signification de 95 % pour un test bilatéral, visant à répondre à la question suivante : la CE 50_1 diffère-t-elle ou non de la CE 50_2 en étant plus petite ou plus grande ? Pour un test unilatéral, visant à répondre à la question suivante : la CE 50_1 est-elle ou n'est-elle pas statistiquement plus grande que la CE 50_2 ?, la signification serait établie si | Z | est plus grand que 1,645.

Autres approches. — Sur cette question, d'autres opinions ont été exprimées relativement à l'écotoxicologie. Villeneuve *et al.* (2000) se sont interrogés sur la puissance relative, ce qui est la même question générale que le fait de déterminer des différences significatives entre les CE 50. Ils ont reconnu que les estimations de la puissance relative (quotient de deux CE 50) ne sont valides que lorsque les droites dose-effet sont parallèles et montrent le même effet réalisable maximal. Dans les essais de toxicité, le respect de la condition de parallélisme est moins important.

Villeneuve *et al.* (2000) ont proposé un cadre d'analyse qui aurait besoin d'être davantage développé pour être utilisé en écotoxicologie. Ils ont exposé une méthode utilisant des estimations en des points multiples dans un intervalle d'effets allant de la CE 20 à la CE 80 pour déterminer les intervalles de puissance relative. Villeneuve *et al.* ont proposé un « cadre » dichotomique à la prise de décisions sur le calcul et à l'application d'estimations de la puissance relative ; toutefois, ils n'ont pas offert de technique mathématique particulière pour traiter les données sur la toxicité. La

63. Des exemples pourraient aider à comprendre. Pour la comparaison des CE 50 dans le tableau, on a fixé arbitrairement toutes les limites de confiance au seuil de 95 % à CE 50 × 1,5 et à CE 50/1,5. Ainsi $f_1 = f_2 = 1{,}5$, et le calcul de $f_{1,2}$ donnera toujours 1,77.

CE 50 max (limites)	CE 50 min (limites)	Quotient des CE 50	$f_{1,2}$	Différents ?
20 (13,3, 30)	8 (5,3, 12)	Pas de superposition	Non calculé	Oui
Comme ci-dessus	11 (7,3, 16,3)	1,82	1,77	Oui, tout juste
Comme ci-dessus	12 (8, 18)	1,66	1,77	Pas tout à fait
Comme ci-dessus	15 (10, 22,5)	1,33	1,77	Non

transformation de la courbe en ligne droite est proposée par l'emploi du logarithme de la dose ainsi que des probits, des logits ou d'outils logistiques. Ils ont subséquemment utilisé la régression linéaire, mais il renvoient le lecteur à plusieurs modèles linéaires généralisés et à d'autres techniques de régression linéaire qui se trouvent dans les publications.

9.5.2 Comparaison de CE 50 multiples

Il ne faut pas répéter entre toutes les paires possibles d'une liste de CE 50 les tests de comparaison deux à deux que montrent les équations 7 et 8, ce qui entraînerait probablement une erreur α (faux positif). Si on a fixé le niveau de signification à 5 %, la répétition du test serait susceptible de conclure à une différence significative, du seul fait du hasard, dans une comparaison sur 20.

Le problème est analogue à celui de l'utilisation répétée d'un test *t* dans les situations où une analyse de variance conviendrait.

En écotoxicologie, on a rarement testé les différences dans une série de CE 50, probablement faute d'une méthode ou d'un progiciel commodes. L'emploi de l'analyse de variance classique pour le test serait valide si on disposait de répétitions des CE 50, mais ce ne serait pas le cas dans la plupart des programmes d'essais.

La méthode décrite par l'équation 9 pourrait servir si on en avait prouvé la validité et si les procédures étaient décrites. L'équation 9 reste provisoire au moment d'écrire ces lignes, mais elle pourrait être développée (Zajdlik, en préparation). La méthode est fondée sur le test du khi-deux et elle permettrait de déterminer si une différence significative ou non existait dans un tableau de plus de deux CE 50. Comme dans l'analyse de variance, l'emploi de l'équation 9 ne permettrait pas de distinguer quelle concentration diffère des autres.

$$\chi^2 = \sum_{i=1}^{k} w_i \left(\log(CE\,50_i) - \left(\frac{\sum_{i=1}^{k} w_i \log(CE\,50_i)}{\sum_{i=1}^{k} w_i} \right) \right)^2 \qquad (9)$$

Pour chaque CE 50, $w = (1/s_{\bar{x}.\log(\text{CE }50)})^2$, c'est-à-dire le carré de la réciproque de l'erreur type ($s_{\bar{x}}$) du logarithme de la CE 50. Certaines étapes du calcul sont exposées dans l'annexe Q, avec un exemple. On pourrait utiliser une feuille de calcul ou un simple programme informatique pour faciliter ce calcul.

On compare le khi-deux (χ^2) calculé aux valeurs de la table pour le nombre de CE 50 moins un et à la valeur choisie de la probabilité, d'habitude p = 0,05. Si la valeur calculée excède celle de la table, il existe au moins une différence significative entre les CE 50.

Pour déterminer quelle(s) CE 50 diffère(nt) des autres, il faudrait un test de comparaisons multiples, mais on n'en a pas encore trouvé un de convenable.

9.6 Différences significatives entre les CI p

Repères

- *On peut comparer deux à deux les CI p par une méthode inspirée du test Z à deux échantillons.*
- *Si plusieurs CI p étaient répétés, on pourrait estimer des différences significatives par l'analyse ordinaire de variance et des tests de comparaisons multiples.*
- *S'il n'y a pas de répétitions, aucune méthode ne permet actuellement de tester les différences entre plusieurs CI p.*

Dans le § 9.5, on expose des méthodes pour tester les différences significatives entre deux ou plusieurs paramètres de toxicité létale et d'autres formes de toxicité quantique. On peut utiliser des méthodes analogues pour estimer les paramètres quantitatifs de toxicité. Certaines méthodes sont établies pour des paires de paramètres de toxicité, mais il n'existe pas encore de méthodes de comparaison de plusieurs paramètres de toxicité.

En Amérique du Nord, le paramètre quantitatif de toxicité le plus utilisé est la CI 25. Comme les méthodes qui suivent sont valables pour n'importe quelle valeur de *p,* on parlera de la CI p. Il est entendu que les comparaisons ne doivent être faites que pour les mêmes valeurs de p, c'est-à-dire une CI 20 avec une

autre CI 20, une CI 25 avec une autre CI 25, etc.

9.6.1 Paires de concentrations inhibitrices (CI p)

Aucune superposition des limites de confiance. — Si les intervalles de confiance des CI p ne se superposent pas, on peut affirmer que ces dernières sont significativement différentes, sans autre forme de procès. Si les intervalles se superposent, cela ne dit rien au sujet d'une différence significative. Le principe est le même que dans la comparaison se deux CE 50 (§ 9.5.1).

Méthode de Litchfield-Wilcoxon. — Cette méthode (§ 9.5.1) utilise une combinaison de limites de confiance de deux CE 50 pour juger du caractère significatif des différences. Bien qu'elle semble facile à étendre aux CI p, les statisticiens s'accordent à dire qu'elle ne convient pas à cet usage.

Méthode particulière nº 2 de Zajdlik. — Cette méthode est semblable à la nº 1, exposée dans le § 9.5.1. Elle découle aussi du test Z à deux échantillons, décrit dans la plupart des manuels de statistique (par ex. Zar, 1974, p. 105-106). L'équation 10 en donne la formule, mais les étapes du calcul restent à décrire (Zajdlik, en préparation). Les manipulations mathématiques sont assez simples, comme le montre l'équation 10, et elles ne comportent que les valeurs logarithmiques des CI p et de leur erreur type. Cette méthode suppose que les CI p ou — plutôt, pour les essais d'écotoxicité — que les logarithmes des CI p obéissent à une loi normale.

$$Z = \frac{\log CI\ p_1 - \log CI\ p_2}{\sqrt{(\log s_{\bar{x}\,CI\,p1})^2 + (\log s_{\bar{x}\,CI\,p2})^2}} \qquad (10)$$

Si | Z | > 1,96 (c'est-à-dire plus grand que la valeur critique de Z), les deux CI p diffèrent significativement au niveau de signification de 95 %, pour un test bilatéral. Dans le cas plus habituel, il serait évident pour l'expérimentateur qu'une des CI p serait numériquement plus grande que l'autre. Un test unilatéral serait approprié (la CI p_1 n'est-elle pas, statistiquement, plus grande que la CI p_2 ?). Une différence statistique serait établie si | Z | était plus grand que la valeur critique de 1,645.

Dans l'équation 10, $s_{\bar{x}}$ représente l'erreur type du logarithme de la CI p. L'erreur type de chaque CI p se calcule de la façon indiquée dans l'équation 11 (Zajdlik, en préparation).

$$S_{\bar{x}(\log CI\,p1)} = e^{(\ln(\log LSC - \log CI\,p_1) + \ln(\log CI\,p_1 - \log LIC) - 2\ln(z_{1-\alpha/2})) \div 2} \qquad (11)$$

Dans cette équation,

LSC est la limite supérieure de confiance de la CI p pour (1 – α) % (habituellement 95 %) ;

LIC est la limite inférieure de confiance de la CI p pour (1 – α) % (habituellement 95 %) ;

Z est le quantile normal pour (1 – α) %. Le quantile est la *LSC* (95 %) moins la *LIC* (95 %) = 1,96, et on introduit cette valeur numérique dans la formule.

L'approche se fonde sur les intervalles de confiance calculés en même temps que les CI p. La méthode est appropriée, que la CI p et ses limites aient été obtenues par régression ou par interpolation et la méthode *bootstrap*. L'équation 11 utilise les limites supérieure et inférieure de confiance pour donner deux estimations de la variance d'une CI p donnée, en l'occurrence la CI p_1, la première CI p. La moyenne de ces estimations permet d'obtenir une seule valeur pour la première erreur type. L'erreur type de la deuxième CI p serait estimée de la même manière.

Les logarithmes népériens de l'équation 11 entrent dans le calcul de la moyenne géométrique des limites supérieure et inférieure de confiance. Les logarithmes (de base 10) représentent les calculs des expositions expérimentales à partir d'une suite logarithmique de concentrations (§ 2.3).

Dans certaines méthodes de calcul de la CI p, notamment dans les meilleures méthodes de régression, l'erreur type ferait partie du résultat de l'analyse. L'expérimentateur pourrait utiliser cette valeur sans

devoir employer l'équation 11 et passer à la comparaison par l'équation 10.

On peut donner un exemple de calcul de l'erreur type par l'équation 11, en utilisant des valeurs arbitrairement choisies pour la CI p (10 mg/L) et ses limites de confiance (6 et 16 mg/L). Dans le deuxième membre de l'équation 11, la partie mise entre parenthèses devient (après omission de quelques chiffres) :

[Ln (log 16 – log 10) + ln (log 10 – log 6) – 2 (ln 1,96)]/2

[– 1,5890... – 1,5057... – 1,3458...] / 2 = – 2,2203...

En prenant cette valeur comme exposant, cela donnerait, pour l'équation :

$$s_{\bar{x}\ (\log CI\ p)} = 0{,}108571336$$

La somme (CI p + erreur type) se calculerait comme suit :

$$\log CI\ p + s_{\bar{x}\ (\log\ CI\ p)} = 1 + 0{,}10857... = 1{,}10857,$$ dont l'antilogarithme est 12,8 mg/L

La différence (CI p – erreur type) se calculerait comme suit :

$$\log CI\ p - s_{\bar{x}\ (\log\ CI\ p)} = 1 - 0{,}10857... = 0{,}89142,$$ dont l'antilogarithme est 7,79 mg/L

9.6.2 Comparaison de CI p multiples

Si, dans plusieurs ensembles d'essais, il y a une véritable répétition de CI p, on peut tester les différences entre les ensembles par les méthodes ordinaires d'analyse de variance, suivies d'un test de comparaisons multiples, si on le désire. Cependant, s'il se trouve une série de CI p non répétées, il ne semble y avoir de méthode en usage pour établir si, entre ces concentrations, une ou des différences sont significatives. Un test de comparaisons deux à deux tel que celui que montre l'équation 10 ne doit pas être répété entre toutes les paires possibles d'une liste de CE 50 en raison du risque de faux positif (erreur α).

La méthode montrée dans le § 9.5.2, pour comparer plusieurs CE 50, semblerait se prêter à la comparaison de CI p et elle pourrait être raffinée à cette fin (Zajdlik, en préparation). Sinon, les efforts pourraient se concentrer sur l'obtention des résultats bruts des essais et l'application à ces résultats de techniques plus perfectionnées.

Quand les résultats sont « difficiles »

Les essais de toxicité peuvent donner diverses formes de résultats, qui les rendent difficiles à traiter. La présente section porte sur certaines difficultés, la plupart concernant les essais de toxicité sublétale, mais toutes ne bénéficient pas de solutions admises.

10.1 Variabilité

Repères

- *La forte variabilité des effets ne devrait pas influer systématiquement sur la CI p, en la poussant vers le haut ou vers le bas, mais un intervalle de confiance plus large signifiera que l'estimation est moins fiable. Dans le test d'hypothèse(s), une grande variabilité déplace les valeurs de la CSEO et de la CEMO vers le haut.*

Si on se sert d'un modèle linéaire pour estimer un paramètre de toxicité, la variabilité, si elle est grande, pourrait ne pas changer la valeur de ce paramètre, bien qu'elle élargisse l'intervalle de confiance. Comme les limites de confiance auront été précisées, la fiabilité du paramètre de toxicité sera manifeste pour tous les utilisateurs de cette information.

Si le test d'hypothèse(s) sert à analyser les résultats d'un essai, il sera rendu moins sensible par la grande variabilité de ces derniers. Le paramètre de toxicité se situera à une concentration plus forte, défaut de l'approche fondée sur le test d'hypothèse, qui est à la base des discussions des § 7.1 ainsi que 7.2.2 à 7.2.5.

À la fin de l'essai de toxicité, la variabilité est fixée. La seule façon d'en réduire au minimum les effets consiste à choisir la méthode d'analyse statistique la plus appropriée et la plus efficace. Si on effectuait des nouveaux essais semblables, le remède le plus probable consisterait, au stade du plan d'expérience, à augmenter la taille des échantillons ou, parfois, à affiner le plan d'expérience statistique et à supprimer ou à réduire les causes de variation dans le mode opératoire.

10.2 Observations aberrantes

De temps en temps, les résultats renferment une *observation aberrante,* qui ne semble pas en harmonie avec les autres résultats de l'essai. L'expérimentateur remarquerait probablement l'observation aberrante d'abord à la lecture des tableaux ou à l'examen du tracé de la distribution des résultats, raison pour laquelle nous insistons pour que l'on trace d'abord à la main un graphique des résultats.

Il n'existe pas de méthode fondée sur les mathématiques ou le jugement qui puisse, de façon magique et définitive, séparer une erreur d'une variation inhérente. L'erreur et la variation pourraient se ressembler dans leur grandeur, et l'expérimentateur ne doit pas céder à la tentation de se débarrasser arbitrairement d'un point qui ne semble pas en harmonie avec une présumée distribution. D'autre part, on ne devrait pas aveuglément traiter un point discordant de la façon habituelle — ce point pourrait, en effet, être erroné et avoir une mauvaise influence sur les interprétations techniques.

Repères

- *Si on remarque une observation apparemment aberrante, il faut une bonne raison pour la supprimer.*
- *Si une observation est aberrante, il faudrait examiner tous les procès verbaux de l'essai, à la recherche d'une erreur humaine. Il faudrait chercher dans les méthodes de garde ou de maintien ainsi que d'essai les causes possibles d'une réaction biologique altérée. On devrait envisager des modèles d'analyse de rechange, peut-être une simple transformation des données.*
- *L'expérimentateur devrait également appliquer des tests mathématiques appropriés pour évaluer les observations aberrantes, comme il est décrit dans le texte. Cependant, ces tests ont des carences pour ce qui concerne leur emploi en*

toxicologie, et on devrait en nuancer les conclusions, grâce à l'examen de la variation totale dans une expérience.

- *On devrait signaler les anomalies ainsi que tout test effectué à leur égard et toute conclusion formulée quant à leur nature.*
- *En général, on devrait analyser les essais avec et sans la valeur aberrante et on devrait, dans les deux cas, signaler les résultats de l'analyse en indiquant ce que l'on considère comme définitif et en motivant cette conclusion.*

On peut tenter de résoudre rationnellement le problème des observations aberrantes. Trois étapes sont à suivre à l'égard des observations suspectes, selon Grubbs (1969) par l'entremise de Newman (1995). C'est un bon conseil pour les écotoxicologues.

(1) L'expérimentateur devrait rejeter toute mesure qu'il sait avoir été obtenue par une méthode défectueuse. Il devrait la rejeter, qu'elle *semble ou non* en harmonie avec la présumée distribution. (On devrait examiner, dans le cadre du programme normal de maîtrise de la qualité du laboratoire, les méthodes influant sur toutes les données, que ces dernières semblent ou non inhabituelles.)

(2) Ensuite, l'expérimentateur devrait envisager la possibilité d'avoir adopté un modèle qui ne convient pas, qui pourrait être la cause du manque d'ajustement d'une ou de plusieurs observations. Cette possibilité, souvent négligée, est importante.

(3) Enfin, si l'anomalie reste inexpliquée, on devrait la signaler, quel que soit le parti choisi pour l'analyse subséquente des données.

10.2.1 Vérification des erreurs et des modes opératoires

Toute observation aberrante devrait faire l'objet d'une nouvelle vérification pour trouver la trace d'une erreur humaine. Cela comprend la mesure de l'effet, l'enregistrement des données, le transfert des chiffres ou leur saisie dans les programmes informatiques.

La solution la plus heureuse pour remédier à une donnée aberrante serait de découvrir qu'elle a été causée à la faveur d'une erreur de transcription ou d'arithmétique, que l'on peut corriger immédiatement. *On devrait vérifier de même tous les autres points de données.* Il pourrait également y avoir une erreur dans une observation non aberrante, et l'on doit soumettre toutes les observations au même examen, au moyen d'une approche scientifique équilibrée.

Si aucun lapsus (de la plume ou du clavier) n'est apparent, l'expérimentateur devrait rechercher les causes biologiques ou celles qui, dans le mode opératoire, pourraient être à l'origine de l'anomalie apparente. Conformément au principe que l'organisme en expérience ne ment jamais, l'expérimentateur devrait envisager tous les stimuli possibles dus à l'environnement et auxquels les organismes ont été soumis pendant l'acclimatation et l'expérience.

Il faudrait examiner toute la suite des modes opératoires dans tous les éléments de l'essai et dans tous les traitements utilisés. Cela suit la première étape susmentionnée.

10.2.2 Modèles de rechange

Si aucune erreur n'est manifeste, l'étape suivante pourrait consister à s'interroger sur la justesse du modèle utilisé. Par exemple, on pourrait poser, par hypothèse, une diminution régulière des performances en raison de l'augmentation de la concentration, mais, en réalité, on pourrait avoir affaire à un phénomène d'hormèse (augmentation des performances à une faible concentration ; § 10.3).

Une autre étape logique dans la recherche d'un modèle plus approprié serait la possibilité de transformer les données au moyen d'une opération courante. Une tendance systématique des données pourrait se prêter à une transformation avantageuse. Par exemple, on pourrait remédier à l'absence générale de normalité dans la distribution des données par la transformation arc sinus (racine carrée). Si l'observation aberrante était un point unique, la justification de la transformation serait moins convaincante, tout comme la probabilité de résoudre la difficulté par cette méthode.

Si la transformation est de peu de secours, l'analyse par des méthodes non paramétriques pourrait être utile. Une méthode de classement peut donner de bons résultats

avec une valeur aberrante, puisqu'elle est habituellement moins exposée à l'influence d'une observation aberrante. L'OCDE (OECD, 2004) propose d'inclure une telle analyse non paramétrique (y compris l'observation aberrante) comme dernière étape supplémentaire dans un rapport qui comprend deux analyses paramétriques (avec et sans l'observation aberrante ; v. le texte qui suit).

Un statisticien pourrait proposer un *modèle robuste*, qui utilise une *fonction de pénalité* différente (une règle d'optimisation telle que les sommes résiduelles minimales des carrés) qui réduit au minimum l'effet de l'observation aberrante. Une comparaison des inférences obtenues grâce à des méthodes ordinaires et robustes pourrait orienter les décisions à prendre sur l'observation aberrante.

10.2.3 Critères applicables aux observations aberrantes

Parallèlement aux méthodes exposées dans les § 10.2.1 et 10.2.2, l'expérimentateur devrait, si possible, utiliser des techniques mathématiques objectives pour voir si l'observation aberrante semble représenter une anomalie ou si elle est simplement une variation. La recherche de ces techniques mathématiques doit être assujettie au bon sens, lorsqu'elle est appliquée à des résultats de toxicité (v. le texte qui suit), mais elle peut aider à décider si la valeur anormale doit ou ne doit pas être incluse dans l'analyse d'ensemble des résultats.

S'il n'y a aucune répétition (comme sur la droite des probits d'un essai quantique, il n'existe aucun moyen objectif de reconnaître les observations aberrantes. Dans son rapport, l'expérimentateur devrait faire connaître l'amplitude de l'anomalie au moyen de données présentées dans des tableaux ou dans des graphiques.

Si on possède des répétitions, on peut prendre des options supplémentaires. La fig. 22 montre des exemples d'observations peut-être aberrantes qui correspondent aux deuxièmes concentrations les plus faibles : dans le graphique de gauche, une valeur semble particulièrement différente des autres.

Règles empiriques. — Pour évaluer une observation aberrante parmi des mesures répétées, on pourrait utiliser une règle empirique. Si l'observation est éloignée de la médiane de plus de 1,5 fois l'intervalle interquartile, elle est probablement aberrante [64]. Malheureusement, la règle empirique perd une certaine utilité en écotoxicologie parce que, habituellement, on ne compte que quelques répétitions par traitement et que les estimations de l'intervalle interquartile deviennent plutôt incertaines. Par exemple, cette méthode n'est pas utile dans le cas des données de la fig. 22, parce qu'il serait chimérique de vouloir estimer des quartiles pour une série de quatre mesures.

Une variante de cette méthode informelle est la règle de Tukey (Tukey, 1977), qui englobe des observations quelque peu moins nombreuses qui seraient des observations aberrantes possibles. Une observation aberrante possible serait inférieure ou supérieure à 1,5 fois l'intervalle interquartile au premier quartile ou au troisième quartile, respectivement. De 1,5 à 3,0 fois l'intervalle susmentionné et on parle d'une observation « légèrement aberrante », tandis que plus de 3,0 fois l'intervalle et on parle d'une observation « gravement aberrante ». Cette méthode éprouve la même difficulté à décider de l'intervalle interquartile de répétitions peu nombreuses comme on les trouve habituellement en écotoxicologie. L'OCDE (OECD, 2004) propose la règle de Tukey comme test formel d'estimation des observations aberrantes en fonction des résidus (différences entre la moyenne du traitement et les valeurs individuelles), afin de ne pas confondre les observations aberrantes et les effets du traitement.

Critères statistiques pour les observations aberrantes. — Des tests statistiques ont été proposés pour évaluer objectivement une éventuelle observation aberrante. Newman (1995) recommande la méthode de Grubbs (1969). On saisit la valeur de l'observation que l'on soupçonne d'être aberrante et qui fait partie d'un groupe dans une formule comprenant la moyenne ($\bar{x}$) et l'écart type (s) de *toutes les observations* pour estimer une valeur T. En écotoxicologie, « toutes les observations » signifie toutes les valeurs obtenues à la concentration particulière à laquelle correspond la

64. *Intervalle interquartile* est défini dans le glossaire et décrit dans l'annexe R. Si 20, 24, 28, 34 et 40 étaient les cinq moyennes d'une série, le premier quartile serait 24, la médiane 28, et le troisième quartile serait 34. L'intervalle interquartile serait de 34 – 24 = 10. Le critère serait 10 fois 1,5 = 15. Les limites seraient de 28 ± 15 = 13 et 43. Les valeurs minimale et maximale de la série se trouvent à l'intérieur de ces limites et ne sont probablement pas des observations aberrantes.

valeur aberrante. Les formules des observations aberrantes supérieures et inférieures à la moyenne sont respectivement comme suit :

$$T = \frac{observation\ aberrante - \bar{x}}{écart\ type}$$

$$T = \frac{\bar{x} - observation\ aberrante}{écart\ type}$$

On compare la valeur calculée de *T* aux valeurs critiques d'une table fournie par Grubbs (1969) et Newman (1995). Si la valeur calculée excède la valeur critique, l'observation que l'on soupçonne d'être aberrante est estimée ne pas provenir de la même distribution normale que le reste des valeurs [65]. Une carence importante de ce test paramétrique de détection des observations aberrante est l'hypothèse selon laquelle les données obéissent à une distribution particulière, normale dans le cas qui nous occupe. Le rejet ou non d'un point de données dépend de cette hypothèse.

On peut appliquer cette formule aux données correspondant à la deuxième concentration la plus faible du graphique de gauche de la fig. 22, où semble se trouver une observation sans conteste aberrante. Les poids moyens des larves de tête-de-boule des quatre répétitions sont : 0,69, 0,77, 0,79 et 1,47 mg (ce dernier étant déviant). Avec des chiffres significatifs supplémentaires pour le calcul, la moyenne est de 0,93, l'écart type est de 0,3626, et la valeur calculée de *T* est 1,49. La valeur critique (v. la note 65) des quatre mesures est 1,46. *T* y est à peine supérieur, de sorte que le point de valeur élevée pourrait être classé comme observation aberrante. Cela semble justifié par l'examen de tout l'ensemble de données. Il est remarquable que *T* ne semble pas excéder la valeur critique autant qu'on pourrait s'y attendre d'après l'aspect du graphique.

Si l'on répète l'opération pour les données du graphique de droite de la fig. 22, on constate que ce test objectif doit être appliqué avec jugement aux données toxicologiques. Les quatre poids moyens des poissons correspondant à la deuxième concentration la plus faible sont 0,84, 0,82, 0,85 et 1,0 mg. La moyenne, l'écart type et *T* sont respectivement de 0,8775, de 0,08261 et de 1,48. Encore une fois, la valeur critique est 1,46. La valeur calculée de *T* excède à peine la valeur critique, de sorte que cela justifie dans une certaine mesure le rejet de la valeur maximale correspondant à cette concentration et l'analyse à l'aide des trois autres valeurs. Cependant, il faudrait examiner toute la distribution des données dans le graphique de droite de la fig. 22. La variabilité globale de la deuxième concentration la plus faible ne diffère pas beaucoup de celle des autres concentrations. La décision statistique du caractère aberrant de l'observation semble avoir été pilotée par le groupage dense des trois autres mesures, qui sont en effet exceptionnellement proches les unes des autres (0,84, 0,82 et 0,85). Cette densité réduit l'écart type à une valeur très faible et, ainsi, elle augmente la valeur calculée de *T*. La distribution que l'on suppose normale pourrait ne pas être valable, problème évoqué précédemment. En outre, cette méthode statistique ne tient pas compte de la variation globale révélée par la totalité de l'essai ; elle subit complètement l'influence de la variation minime à la concentration à laquelle on s'intéresse, qui était apparemment un événement fortuit peu commun. Une méthode qui intégrerait la variation totale dans un essai répondrait mieux aux besoins de l'écotoxicologie.

Dans un cas limite comme celui qu'illustre le graphique de droite de la fig. 22, la prudence dicterait l'acceptation de la mesure mise en doute. Il est conseillé à l'expérimentateur de produire les résultats des analyses avec et sans le point limite, avec une description de la situation et des conclusions interprétatives.

D'autres méthodes statistiques permettant de déceler les observations aberrantes sont décrites dans les publications, mais elles semblent posséder les mêmes

65. Dans la table de Grubbs (1969), on trouve un choix de trois niveaux de signification pour jusqu'à 100 observations dans la distribution. En écotoxicologie, le nombre inférieur d'observations serait la règle. Pour 3, 4, 5... 10 observations dans la distribution et un niveau de signification de 5 %, les valeurs critiques seraient de 1,15, 1,46, 1,67, 1,82, 1,94, 2,03, 2,11 et 2,18. La valeur critique pour 20 et 30 observations serait respectivement de 2,56 et de 2,75. Il s'agit de valeurs unilatérales, comme il serait approprié pour les formules montrées.

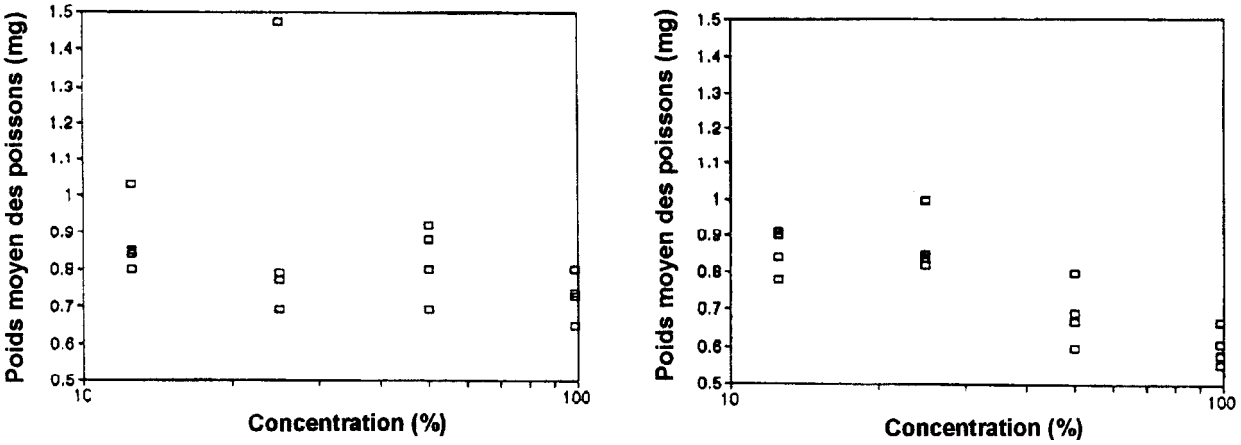


Figure 22. — Exemples d'observations peut-être aberrantes dans des essais de mesure de la croissance, au 7e jour, de larves de têtes-de-boule. Ces données proviennent d'essais sur deux effluents de fabriques canadiennes de pâte à papier. Dans chaque exemple, une mesure à la 2e concentration la plus faible (25 % d'effluent) excède les autres mesures à cette concentration et, aussi, la distribution générale. Le nombre de répétitions à chaque concentration était de quatre, chacune avec 9 ou 10 larves.

faiblesses à l'égard des essais de toxicité, c'est-à-dire de ne pas prendre en considération la variation totale de toutes les concentrations. Le manuel classique de statistique de Snedecor et Cochran (1980) offre deux formules relativement simples permettant de tester la nature aberrante ou non d'une observation. Les méthodes reposent sur l'emploi de tables assez détaillées des valeurs critiques des tests, que nous ne pouvons pas reproduire ici. Une monographie a été exclusivement consacrée à la question par Barnett et Lewis (1994), et l'USEPA (1995) conseille la consultation d'une publication de Draper et John (1981).

Observations aberrantes multiples. — L'application d'un remède a une issue plus douteuse lorsque, à une concentration donnée, on soupçonne plus d'une observation aberrante. Collett (1991) était d'avis qu'il n'existe pas de méthode objective et fiable que l'on peut recommander pour évaluer un groupe de deux ou de plusieurs observations aberrantes ; cependant, Rosner (1983) offre une méthode exploitable. La même méthode est décrite par Newman (1995), avec un code de programme informatique en FORTRAN. Snedecor et Cochran (1980) ont montré comment les deux formules simples concernant une observation aberrante pouvaient et devraient être appliquées au cas de deux valeurs aberrantes dans un ensemble d'observations (disons dans les répétitions correspondant à une concentration). Il faudrait tester d'abord la valeur aberrante la plus extrême. Que cette valeur *se révèle ou non* une observation aberrante statistique, on devrait la supprimer et tester l'autre valeur extrême dans la distribution résiduelle de valeurs. Si cette valeur se révèle statistiquement aberrante, les deux sont déclarées aberrantes. Cette opération se justifie comme suit : la valeur la plus extrême peut « occulter » l'écart que représente la seconde valeur extrême, en influant sur l'ensemble de la distribution.

Bien qu'il n'existe pas de critère tout à fait adéquat pour les observations aberrantes dans les essais de toxicité, la méthode de Grubbs (1969), que nous avons exposée ci-dessus, semble aussi appropriée que toute autre. Il faudrait mettre au point des méthodes statistiques convenant davantage à la toxicologie. En attendant, si une observation apparemment aberrante est cruciale pour l'interprétation d'un essai, d'une façon ou d'une autre, on devrait consulter un statisticien pour appliquer des mesures, que nous ne décrirons pas ici, afin de quantifier le degré d'influence qu'une observation particulière exerce sur un modèle.

10.2.4 Interventions à signaler

Si on soupçonne une observation d'être aberrante, on peut résumer la séquence des approches et des

interventions souhaitables comme suit :

- On devrait examiner, à la recherche d'erreurs dans les observations ou les enregistrements, tous les enregistrements de l'essai.
- Ensuite, on devrait examiner toutes les méthodes utilisées dans le maintien des sujets de l'expérience et dans leur exposition aux conditions expérimentales, pour voir si elles n'ont pas déclenché quelque réaction biologique compréhensible.
- Sinon, on devrait envisager d'autres modèles de rechange pour les résultats.

Parallèlement à ces étapes, l'expérimentateur devrait utiliser des méthodes statistiques objectives pour examiner la question du rejet ou de l'acceptation de l'observation ou des observations variantes.

Dans le rapport, on devrait énumérer les anomalies et décrire les résultats et les conclusions des recherches afférentes.

Si, *selon* les techniques statistiques, l'observation *n'est pas* aberrante, on devrait le signaler, et l'analyse de l'ensemble des résultats devrait englober le résultat aberrant. (On pourrait également inclure dans le rapport une analyse sans la valeur anormale, accompagnée d'observations sur les conséquences que cela a eues sur l'interprétation.)

On devrait également signaler si le test statistique a permis de reconnaître une observation aberrante. On devrait analyser les résultats avec et sans la valeur douteuse. Les deux analyses devraient être mentionnées, l'expérimentateur devrait préciser laquelle il a choisi comme définitive, en motivant son choix. Une analyse différente ou supplémentaire, par une méthode non paramétrique ou une autre méthode plus robuste, pourrait projeter un éclairage supplémentaire. Cette répétition des analyses et des explications pourrait ne pas convenir à certains programmes réglementaires, qui exigent habituellement un résultat normalisé. Dans ce cas l'expérimentateur devrait communiquer la meilleure estimation, selon son jugement, en ajoutant que des analyses et des explications supplémentaires du contexte sont jointes ou archivées.

10.3 L'hormèse — stimulation à faibles concentrations

Repères

- *Dans beaucoup d'essais de toxicité sublétale, les performances des sujets d'expérience sont stimulées à de faibles concentrations. C'est le phénomène de l'hormèse (les sujets font mieux que les témoins). Cela pose le problème philosophique consistant à décider si ces effets sont nuisibles et quelles performances devraient être considérées comme celles du témoin. On ne peut pas donner de réponse générale à cette question.*
- *L'hormèse pose également des problèmes pratiques d'analyse. Les modèles ordinaires dose-effet ne sont pas satisfaisants ou ils faussent les estimations. Les modèles plus complexes peuvent perdre leur pouvoir de détection des concentrations nuisibles, si on adopte un plan minimal d'expérience.*
- *Pour les estimations ponctuelles, telles que celles de la CI 25, la meilleure approche consiste à ajuster les données avec une régression non linéaire, puis à estimer la CI 25 par comparaison avec le vrai témoin. Dans les nouvelles méthodes d'Environnement Canada et dans le présent document, on présente une approche analytique normalisée de ce phénomène.*
- *Si on effectue un test d'hypothèse, on devrait suivre le mode opératoire normal, avec les résultats obtenus à toutes les concentrations. Cependant, on ne devrait prendre en considération que les effets significativement pires que ceux du vrai témoin dans la désignation de la CSEO et de la CEMO.*
- *À l'égard des essais hormétiques, les rapports devraient comprendre les résultats originels et expliquer les méthodes d'analyse.*

L'hormèse est le fait, pour une matière, de stimuler à de faibles concentrations les organismes en expérience par rapport aux organismes témoins, c'est-à-dire qu'ils « performent mieux » que les témoins. Aux fortes

concentrations, les effets nuisibles sont visibles. Habituellement, la terminologie la plus juste est, en termes plus généraux, la *stimulation à faible dose*. L'expression englobe d'autres causes possibles de stimulation, telles que l'effet de solvants, l'erreur expérimentale ou, en théorie, uniquement une stimulation générale des organismes en expérience gardés dans des conditions monotones de laboratoire (« *stimulation suffisante* »). On perçoit parfois la stimulation à faible dose dans divers effets, notamment l'accélération de la croissance des organismes ou l'augmentation de la densité des cellules algales, montrée dans la fig. 23.

La stimulation à faible dose est peut-être l'empêchement le plus fréquent de l'analyse des résultats d'essais de toxicité sublétale bien planifiés. Elle représente un phénomène réel et non des observations aberrantes ou un essai imparfait, et des expérimentateurs canadiens tombent sur des résidus qui donnent de manière fiable des résultats hormétiques. Cela ne se limite pas aux études de l'environnement ; le phénomène est répandu en toxicologie médicale (Davis et Svendsgaard, 1990). Calabrese et Baldwin (1997) ont examiné des effets positifs allant de 30 à 60 %, bien que des laboratoires canadiens de l'environnement observent, comme maximum le plus probable, un taux de + 30 %.

La cause de performance accrue est rarement déterminée. Pour ce qui concerne la croissance ou le nombre de cellules produites, les constatations correspondraient à l'apport de nutriments avec la matière à l'étude, ce qui stimulerait la production. Si cela était et si le nutriment était connu, la parade évidente serait d'ajouter le nutriment en question à toutes les concentrations, y compris dans le milieu témoin. Le niveau de compréhension justifierait rarement une telle mesure.

10.3.1 Les difficultés

Problèmes avec les méthodes usuelles. — Si la stimulation à faible dose se manifeste et si on applique des techniques courantes d'analyse, le paramètre de toxicité tend habituellement à être abaissé (« plus grande toxicité »). Les performances du témoin ou les performances de base sont souvent surestimées, ce qui mène souvent à une surestimation des effets des concentrations et à l'abaissement de la CI p. La pente de la relation ajustée devient habituellement plus raide, ce qui pourrait influer sur l'estimation des limites de confiance.

Si les données présentées dans la fig. 23 sont saisies telles quelles dans le programme ICPIN (§ 6.4.3), le processus ordinaire de lissage augmente le nombre de cellules du témoin et il estime une CI p inférieure à celle à laquelle on pourrait s'attendre. La valeur témoin originelle (réelle) de 2 650 cellules est ajustée à 2 860. Le lissage affecte la même valeur (2 860 cellules) à chacune des quatre premières concentrations. En conséquence, si on estime la CI 25, elle est fondée sur 2 145 cellules (75 % de 2 860) plutôt que sur le nombre initial de 1 988. On estime la CI 25 à environ 3,92 (logarithme), soit 8 300 unités de concentration, contre 4,05 ou 11 220 unités de concentration si on avait utilisé le témoin initial, ce qui est assurément un changement appréciable.

Il n'est donc certes pas souhaitable de simplement ignorer la stimulation à faible dose et d'utiliser mécaniquement les méthodes statistiques utilisées communément.

Effet ou témoin plus « performant » ? — Voilà une énigme philosophique que soulève l'hormèse et à laquelle il n'y a pas de réponse consensuelle. Aucun des éléments possibles de solution n'est entièrement satisfaisant.

Les performances améliorées devraient-elles être considérées comme un « effet » du toxique et, par définition, être jugées répréhensibles ? À l'extérieur du laboratoire, dans une communauté vivante, le détournement de l'énergie d'un organisme vers des voies telles que la croissance pourrait en effet être néfaste.

Cela pourrait soustraire de l'énergie qui servirait, plus stratégiquement, à la reproduction ou à quelque autre activité. Cependant, dans un essai en laboratoire, il est difficile d'adhérer à une telle spéculation. Il est difficile de considérer l'accroissement des rendements dans le critère de validité de l'essai comme un effet *nocif*.

D'autre part, si on décidait de considérer la stimulation à faible dose comme un effet nuisible, pour les besoins de l'exemple présenté dans la fig. 23, cela signifierait

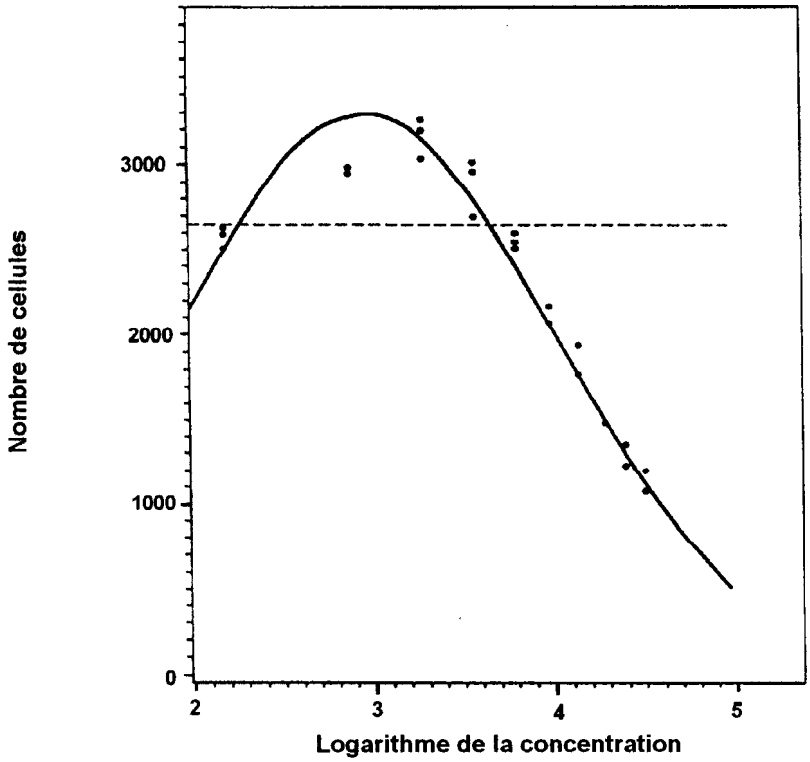


Figure 23. — Exemple de stimulation à faible concentration. Le phénomène a été observé dans un essai avec l'algue verte *Pseudokirchneriella subcapitata* [auparavant *Selenastrum capricornutum*] dans un laboratoire canadien. Le tireté horizontal montre le comportement du témoin, et la concentration minimale a donné lieu à une réponse semblable. L'algue a manifesté un taux de reproduction accru aux trois concentrations plus grandes, puis la baisse prévue aux fortes concentrations.

que les 2e, 3e et 4e concentrations seraient déclarées comme des effets potentiellement toxiques, ce qui ne semblerait probablement pas rationnel à la plupart des observateurs.

Une autre interprétation encore serait de considérer l'augmentation des performances de l'organisme comme une sorte de témoin stimulé, peut-être dû à un apport amélioré de nutriments ou, peut-être, dans le cas des animaux, cela pourrait être une réponse à un « stimulus suffisant » par rapport à un ensemble de conditions par ailleurs monotones d'existence dans les conditions contrôlées du laboratoire. La mesure après stimulation représenterait les performances potentielles et elle jouerait le rôle du témoin. La plupart des expérimentateurs considéreraient probablement cela comme peu réaliste et ils opteraient, plutôt, pour la comparaison avec le témoin ordinaire.

Il n'est pas souhaitable d'adopter un mélange de performances du témoin et de performances après stimulation comme nouveau témoin, tel qu'on l'a fait dans le lissage par le programme ICPIN. Cette méthode, comme cela a été montré, aboutit à un paramètre de toxicité sensiblement abaissé.

La fig. 23 montre la difficulté de répondre à de telles questions fondamentales, et, par le passé, il n'y a pas eu de véritable consensus sur ces questions.

À part les questions philosophiques, la stimulation à faible dose pose des problèmes très pratiques d'élaboration d'une approche statistique pour l'analyse des données (v. les options exposées ci-dessous). Un problème potentiel est que les effets hormétiques à de faibles concentrations signifieraient que la méthode statistique avait moins de concentrations montrant la diminution des performances pour la modélisation de

leur réduction par rapport au témoin.

10.3.2 Prise en considération des effets hormétiques dans la régression

Environnement Canada a adopté cette option raisonnable dans ses méthodes usuelles récentes de régression non linéaire pour estimer les CI p (EC, 2004a, b et 2007 ; § 6.5.8 et annexe O). Pour les comparaisons de l'effet, on utilise le vrai témoin. On utilise un modèle particulier, adapté à l'hormèse et, grâce à lui, on résout les problèmes d'analyse auxquels nous avons fait allusion. Cette solution heureuse de la partie statistique recèle de nombreux avantages. Nul besoin de rejeter aucune des données ; nul besoin de modifier la valeur du témoin par lissage ou par d'autres techniques. Il n'y a pas de distorsion des effets aux concentrations supérieures à celles auxquelles se manifeste l'hormèse. Cela est mentionné sous la rubrique « Option 1 » du § 10.3.3, pour l'analyse des données révélant une stimulation à faible dose.

On a appliqué des modèles linéaires généralisés (GLIM)) aux résultats des essais avec *Ceriodaphnia*. Ces modèles se sont montrés prometteurs pour la stimulation à faible dose, qui s'est souvent manifestée (Bailer *et al.*, 2000a). Les GLIM ont permis des estimations plus cohérentes de la CI p que le programme ICPIN. Ils étaient également applicables aux données quantiques, aux données quantitatives ou aux dénombrements.

Brain et Cousens (1989) ont décrit des ensembles hormétiques de résultats obtenus par des modèles logistiques (sigmoïdes) reparamétrés. Le paramètre ajouté à l'équation a autorisé une modification hormétique des performances à faibles concentrations. On a perfectionné l'approche en incluant le paramètre voulu de toxicité en tant que paramètre (van Ewijk et Hoekstra, 1993), puis on a intégré cette technique dans les méthodes récentes d'Environnement Canada. L'avantage est que l'on estime directement le paramètre de toxicité et ses limites de confiance à partir des données. Un inconvénient possible est la nécessité d'estimer quatre paramètres par régression non linéaire, ce qui exige un plan d'expérience produisant un ensemble de données comportant un nombre adéquat de concentrations et de répétitions. Le paramétrage de van Ewijk et Hoekstra (1993) est sensible à l'algorithme d'optimisation sous-jacent au progiciel non linéaire (B.A. Zajdlik, B. Zajdlik & Associates Inc., Rockwood, Ont., communication personnelle, 2004).

La méthode plus perfectionnée d'analyse supprime la partie statistique de l'énigme posée par l'hormèse. Elle résout les problèmes philosophiques exposés dans le § 10.3.1, grâce à l'approche raisonnable qui consiste à utiliser les performances du vrai témoin comme base pour juger des effets (Option 1, § 10.3.3). La question philosophique, qui est « que devrait-on désigner comme performances "normales" ? », pourrait encore être débattue dans certaines situations inhabituelles, et c'est ce dont il est question dans le § 10.3.3.

10.3.3 Options face à l'hormèse

Dans les options exposées ci-dessous pour les essais dans lesquelles on observe un effet de stimulation à une ou à deux faibles concentrations, on expose une gamme d'approches. Nous recommandons l'option 1, premier choix exigé dans les méthodes récentes d'essai d'un sol d'Environnement Canada (EC, 2004a, b et 2007). Nous recommandons l'option 4 s'il est nécessaire d'obtenir des estimations ponctuelles par le programme ICPIN et l'option 5 pour estimer la CSEO et la CEMO, si le test d'hypothèse(s) est utilisé pour quelque raison que ce soit.

(Option 1) Dans les estimations ponctuelles, inclure l'hormèse dans un modèle plus complexe. — Si on estime la CI 25, adopter le modèle d'hormèse et effectuer une régression non linéaire (§ 6.5.8). La CI 25 est encore estimée relativement aux performances du vrai témoin.

(Option 2) Lissage des effets pour le témoin et les faibles concentrations. — Le lissage est effectué dans le programme informatique communément utilisé ICPIN, pour estimer la CI p. Cela permet d'ajuster le témoin à de « meilleurs niveaux », avec abaissement consécutif de la CI p estimée. On pourrait obtenir un résultat analogue dans le test d'hypothèse(s). Cette option n'est pas recommandée, parce qu'elle fait des comparaisons avec un témoin qui, de fait, n'existe pas.

(Option 3) Omettre de l'analyse statistique les concentrations présentant une hormèse notable. — Cette option ne possède aucune base mathématique ni statistique ; on pourrait la considérer uniquement comme une manifestation du jugement du

biologiste. Cette option permettrait l'ajustement de l'estimation de la CI p par la méthode ICPIN. Elle ne conviendrait pas aux estimations ponctuelles par régression. Cette technique exigerait une analyse préliminaire pour décider quelles concentrations ont effectivement été hormétiques. Une fois les points de données supprimés, l'analyse pourrait porter sur les concentrations restantes. On ferait accompagner l'analyse d'une déclaration claire et motivée des valeurs omises.

Cette option est prise dans une seule des méthodes d'essai de toxicité d'Environnement Canada, celle de l'inhibition de la croissance des algues (EC, 1992d). Si la croissance des algues à une concentration est supérieure à celle du témoin, on signale ces observations, mais celles-ci n'entrent pas dans le calcul de la CI p.

Cette option serait insatisfaisante dans de certaines circonstances. Elle pourrait estimer un paramètre de toxicité qui serait exagérément faible, si la suppression des données hormétiques laissait une large « trouée » entre deux faibles concentrations qui encadraient le paramètre de toxicité. Un exemple hypothétique est décrit dans la ligne 2 de la note ci-dessous [66]. À cause de cette utilité incertaine, la méthode n'est pas recommandée comme solution complète pour les autres essais que ceux de croissance d'algues.

(Option 4) Attribution de la valeur témoin aux concentrations présentant une stimulation à faible dose. — C'est une option arbitraire, sans justification statistique, mais elle est susceptible d'aboutir à l'estimation de paramètres de toxicité réalistes, employant des calculs simples au moyen de méthodes communément utilisées. Nous ne la recommandons pas pour les estimations ponctuelles, parce qu'une méthode appropriée de régression existe (§ 6.5.8). L'option 4 fonctionnerait pour les estimations ponctuelles avec le programme ICPIN. Un échantillon d'expérimentateurs canadiens (Schroeder et Scroggins, 2001) a déjà préconisé cette option pour un usage transitoire, mais seulement jusqu'à ce que des méthodes appropriées de régression aient été mises au point, comme c'est le cas désormais.

(Option 5) Dans un test d'hypothèse(s), considérer la stimulation à faible dose comme non nocive. — L'analyse statistique se déroule comme à l'accoutumée, c'est-à-dire qu'elle engloberait les performances meilleures que celles du témoin. Si elle montrait qu'une ou plusieurs faibles concentrations correspondent significativement à de meilleures performances que celles du témoin, on le signalerait, mais on ne considérerait pas cela comme un effet nuisible. La CEMO serait désignée comme la concentration minimale ayant entraîné une *diminution* significative des performances par rapport à celles du témoin. Le paramètre de toxicité serait le même que celui que l'on aurait estimé dans l'option 4, mais on le préfère pour le test d'hypothèse(s) parce qu'il ne comporte aucune manipulation des données originelles.

Ces options pourraient ne pas convenir à tous les résultats. L'expérimentateur devrait examiner les données portées sur un graphique pour déterminer

66. Cet exemple hypothétique représente le nombre de cellules algales dénombrées à diverses concentrations expérimentales. Dans un souci de simplicité, nous ne reproduisons pas les répétitions.

	Témoin	6	12	25	50	100
				(mg/L)		
1. Nbre observé de cellules	200	200	275	300	100	50
2. Concentrations hormétiques supprimées	200	200			100	50
3. Valeur du témoin à la place de l'hormèse	200	200	200	200	100	50

L'effet observé dans la première ligne dénote des effets hormétiques évidents à 12 et à 25 mg/L. Ces deux effets sont supprimés dans la deuxième ligne, comme dans l'option 3 du texte. Si on voulait estimer la CI 25 (concentration correspondant à 150 cellules) à l'aide de la méthode ICPIN et des données en ligne 2, il y aurait interpolation entre 6 et 50 mg/L, valeurs qui sont séparées par un intervalle plutôt large. On estimerait la CI p à 17 mg/L, ce qui est exagérément faible, puisqu'il n'y avait pas de preuve d'atteinte à la production d'algues dans les données originelles correspondant à 25 mg/L.

La troisième ligne du tableau montre une méthode ayant apparemment été utilisée par certains laboratoires pour contraindre les données à un paramètre de toxicité plus réaliste (option 4 du texte). La valeur du témoin est affectée de façon arbitraire aux concentrations qui se sont révélées hormétiques. L'interpolation de la concentration se situerait désormais entre 25 et 50 mg/L. La CI 25 serait de 35 mg/L, ce qui semble plus acceptable.

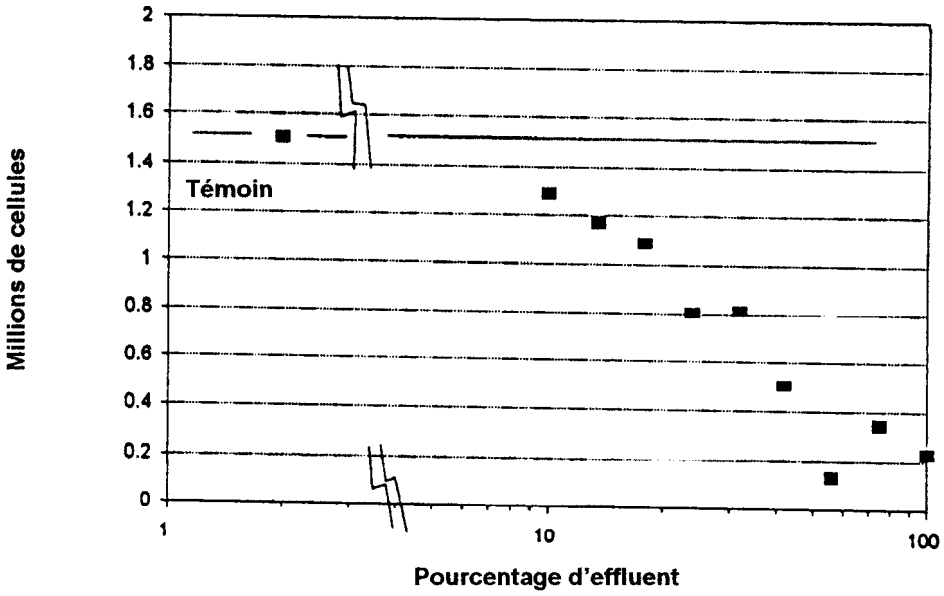


Figure 24. — Exemple de bonne relation linéaire entre la concentration et l'effet. Ce sont les résultats d'un essai sur une eau de surface toxique, effectué au Canada avec l'algue *Pseudokirchneriella subcapitata* [auparavant *Selenastrum capricornutum*]. Pour faciliter la représentation de la concentration sur l'échelle logarithmique, on attribue au témoin une concentration très faible. Le double zigzag figure une interruption dans l'échelle des concentrations.

quelles approches et quels paramètres de toxicité sont acceptables.

Dans tous les cas de stimulation à faible dose, il importe de :

- communiquer les données originelles ;
- préciser les mesures adoptées pour l'analyse.

10.4 Relations concentration-effet déviantes

Repères

- *On donne des exemples de plusieurs types inhabituels ou difficiles d'ensembles de données. On offre des conseils sur leur interprétation.*
- *On peut éviter une partie des difficultés grâce à un plan d'expérience approprié, particulièrement en utilisant un éventail suffisamment étendu de concentrations.*

La plupart des laboratoires tombent parfois sur des relations concentration-effet inhabituelles. Les organismes en expérience mentent rarement, de sorte que les résultats aberrants trouvent d'habitude une explication, laquelle, cependant, pourrait ne pas être évidente. Le présent paragraphe montre des graphiques correspondant à certaines constatations inhabituelles, accompagnées d'explications possibles et de recommandations pour le traitement des résultats. L'interprétation initiale devrait se fonder sur des graphiques, comme il est recommandé dans les § 4.2.2 et 6.3.1 ainsi que dans le guise de l'USEPA (2000a).

La série commence par de « bonnes » données, pour les besoins de la comparaison. Certains exemples anormaux ont été obtenus dans des laboratoires canadiens, pendant des programmes ordinaires d'essai, d'autres s'inspirent d'exemples utilisés par l'USEPA (2000a). Les problèmes posés par les observations aberrantes et l'hormèse ont été abordés dans les paragraphes précédents.

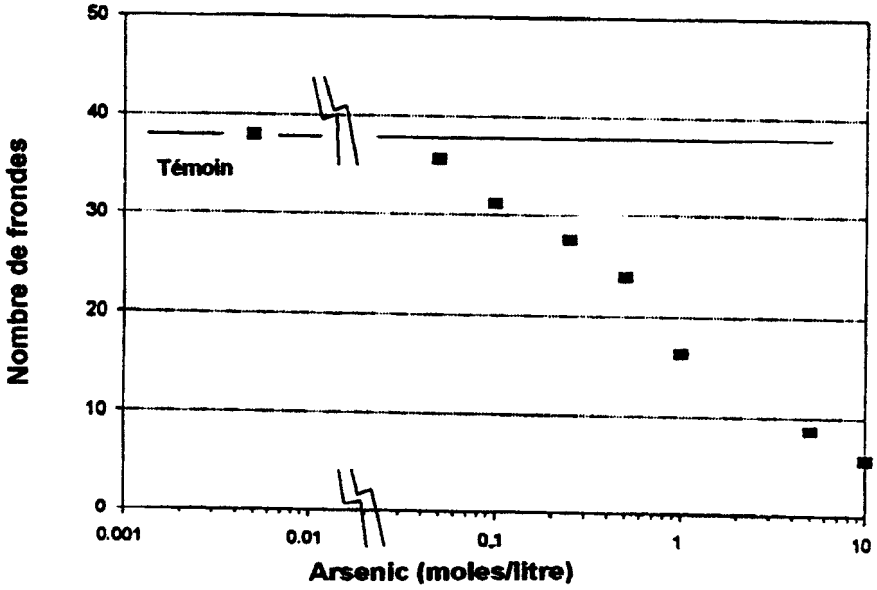


Figure 25. — Autre exemple d'une bonne relation entre la concentration et l'effet. Résultats d'un laboratoire canadien dans un essai de croissance de la lentille d'eau (*Lemna minor*) exposée à diverses concentrations d'arsenic. (Le reste de la description trouvée à la fig. 24 s'applique ici également.)

(1) Bonnes données sur la relation concentration-effet. — La fig. 24 montre une relation linéaire ordinaire pour un essai ayant porté sur des algues. Il N'est aucunement difficile d'estimer un paramètre de toxicité tel que la CI 25 par diverses méthodes linéaires. Le test d'hypothèse(s) fonctionne également d'une manière satisfaisante.

Il serait agréable d'obtenir les bons résultats montrés dans la fig. 24, mais on aurait pu les améliorer dès l'étape du plan d'expérience. La gamme des huit concentrations expérimentales couvre un ordre de grandeur, c'est-à-dire que les concentrations sont rapprochées les unes des autres. Comme nous l'avons mentionné dans le § 2.2, un tel plan d'expérience risque de passer à côté des concentrations intéressantes et, de fait, c'est ce qui s'est produit dans cet exemple. La concentration minimale correspond à un résultat qui est inférieur d'environ 13 % à celui qui correspond au témoin. Une qualité d'un bon plan d'expérience est d'utiliser au moins une faible concentration qui correspondra à des résultats essentiellement analogues à ceux du témoin. Les concentrations auraient dû avoir été étalées sur une étendue plus grande.

Les résultats de la fig. 24 sont généralement monotones, et l'effet légèrement irrégulier de la troisième concentration maximale n'aurait pas de quoi inquiéter. Vraisemblablement, ils représentent la variabilité naturelle et ils contribueraient à augmenter la variance de toute description statistique d'une droite ajustée.

La fig. 25 montre une relation concentration-effet similaire et remarquablement rectiligne. L'analyse de tels résultats ne présenterait aucune difficulté, soit pour une estimation ponctuelle, soit pour un test d'hypothèse(s). Cet étalement des effets sur plus de deux ordres d'ampleur est quelque peu inhabituel. Cependant, cela n'empêcherait pas l'analyse ; le plan d'expérience était adéquat, et on a obtenu de petits et de grands effets. L'addition d'une faible concentration à la série pourrait avoir, cependant, un effet près de celui qu'a éprouvé le témoin. De nouveau, ce bon résultat montre l'importance d'un plan d'expérience englobant une grande étendue de concentrations plutôt que d'essayer de deviner dans quel intervalle étroit se trouvera la concentration importante (§ 2.2). Dans cet essai, l'étalement étonnamment large des effets a englobé toutes les concentrations d'un plan

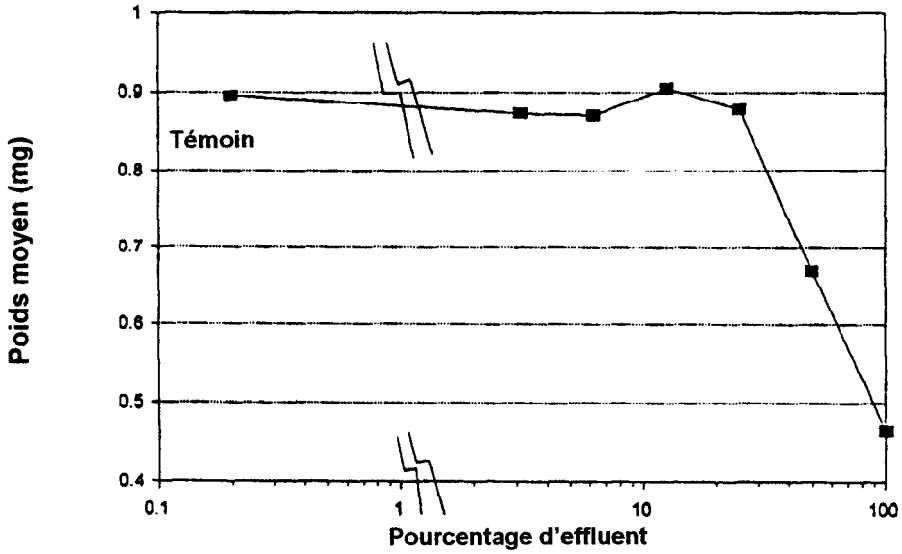


Figure 26. — Relation à pente raide entre le poids des larves de têtes-de-boule et les concentrations d'un effluent auxquelles elles sont exposées. Résultats d'un laboratoire canadien. (Le reste de la description trouvée à la fig. 24 s'applique ici également.)

d'expérience qui, normalement, serait considéré comme d'une grandeur adéquate.

(2) Relations à pente forte. — Dans les essais d'écotoxicité, il est fréquent que, d'une concentration à la suivante, les effets varient brusquement. L'exemple de la fig. 26 n'est pas tout à fait « en tout ou rien » parce qu'il existe un effet intermédiaire quand la relation passe d'une valeur témoin à un effet nuisible majeur. Ce type de données est modérément satisfaisant, et le paramètre estimé de toxicité sera convenablement précis, avec un intervalle étroit de confiance (selon le facteur de dilution utilisé pour choisir les concentrations).

Une qualité évidente de la fig. 26 est la présence d'une faible concentration à laquelle correspond un effet semblable à celui qu'éprouve le témoin. C'est un signe que le plan d'expérience et le mode opératoire de l'essai sont appropriés. On dénombre effectivement quatre faibles concentrations semblables à celles auxquelles le témoin est exposé, et les statisticiens souligneraient l'amélioration de la précision qu'aurait apportée un plus grand nombre de points de données dans la région où l'effet change rapidement. En conséquence, un plan amélioré d'expérience aurait, dans ce cas, omis certaines des faibles concentrations, afin d'obtenir plus de données aux fortes concentrations. Idéalement, un essai mené pour trouver la gamme de concentrations à utiliser aurait montré la suite appropriée de concentrations à utiliser dans l'essai de toxicité définitif. Cependant, faute d'un tel essai préliminaire, toute modification des concentrations prévues dans le plan d'expérience représenterait un jugement après coup. Comme nous l'avons fait remarquer, un plan d'expérience qui rétrécit la gamme de concentrations à utiliser peut être dangereux dans l'essai d'une matière dont on ignore la toxicité. On pourrait passer à côté de concentrations importantes, de sorte qu'il est mieux d'étaler les concentrations, comme cela a été fait.

(3) Absence d'effet et irrégularité. — Parfois, à la concentration maximale, aucun effet de la matière à l'étude n'est évident. Si la matière est un effluent ou un échantillon de sédiment ou de sol, on ne peut pas tester de concentrations supérieures à 100 %. L'interprétation est simple : l'essai n'a révélé aucun effet nocif. On ne peut calculer aucune estimation ponctuelle de la CI p, et le test d'hypothèse(s) montrerait aussi l'absence d'effet.

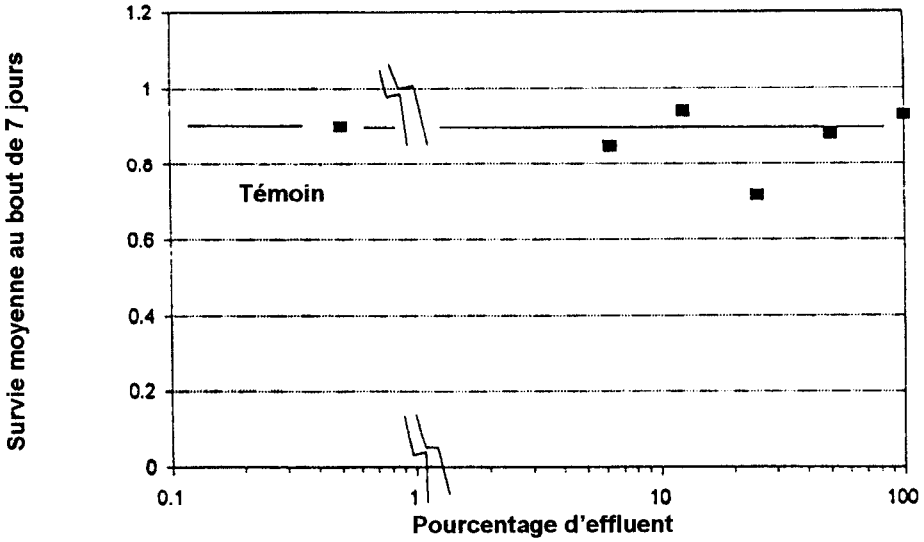


Figure 27. — Absence d'effet aux fortes concentrations avec anomalie à une concentration intermédiaire. Données hypothétiques sur la survie de larves de têtes-de-boule.

C'est ce que montre la fig. 27, mais un résultat révèle une contradiction. La concentration médiane est sensiblement inférieure à celle du témoin. Il est rare qu'un laboratoire puisse observer une telle distribution des résultats dans un essai de toxicité sublétale. À la concentration anormale, les performances des organismes pourraient être diminuées de 25 % par rapport à celles du témoin et pourraient également être statistiquement différentes de celles du témoin.

Si l'analyse des essais se fait normalement par des estimations ponctuelles, cette irrégularité ne pose pas problème. La faible valeur ne se traduirait pas par un paramètre de toxicité, et on devrait la signaler comme une anomalie. Si l'expérimentateur avait l'intention d'utiliser le test d'hypothèse(s), l'effet irrégulier qui correspond à la concentration médiane pourrait se révéler être la CEMO. L'absence d'effet aux concentrations supérieures invalide toute estimation du genre. La seule conclusion raisonnable serait de reconnaître l'anomalie apparente et de déclarer que le test d'hypothèse(s) n'était pas approprié.

On devrait chercher une explication. On examinera les enregistrements pour y déceler des conditions expérimentales divergentes telles que le pH ou l'oxygène dissous. Une seule répétition divergente pourrait avoir influé sur la moyenne (v. le § 10.2 sur les observations aberrantes). Il se peut que l'absence de randomisation ait influé sur les résultats par le truchement de l'état des organismes ou de quelque autre facteur relié à la position dans le tableau de données. Si on ne peut pas trouver d'explications, il reste peu de choses à faire si ce n'est de décrire l'étendue des résultats obtenus et de conclure à l'existence d'un point de donnée anormal.

(4) Absence anormale d'effet à une concentration intermédiaire. — Parfois, une augmentation progressive apparente de l'effet est interrompue par une concentration ne manifestant aucun effet, comme chez le témoin (fig. 28). On pourrait effectuer une analyse par des méthodes qui estiment une CI p. Les techniques d'ajustement des courbes tiendraient compte des irrégularités et produiraient un intervalle de confiance convenablement large. Le programme ICPIN imposerait la monotonie à la relation (Norberg-King, 1993), probablement avec une analyse satisfaisante dans ce cas. Le test d'hypothèse(s) serait gâché par un point anormal significativement différent du témoin ; deux ensembles seraient produits pour la CSEO et la CEMO. L'USEPA (2000a) recommande alors de choisir la valeur inférieure comme CSEO (6,25 % dans la fig. 28) si le test présente une différence significative

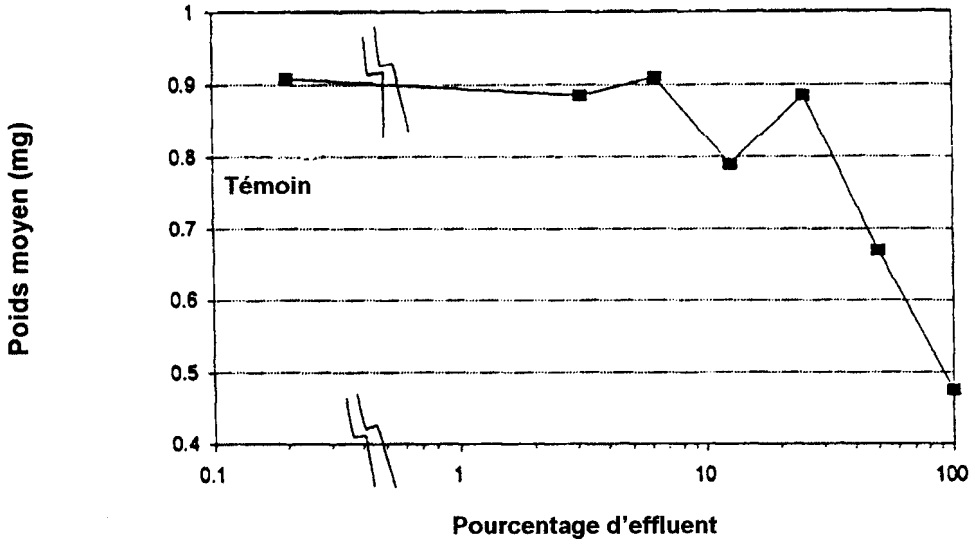


Figure 28. — Absence d'effet, apparemment anormale, à une concentration intermédiaire. Exemple hypothétique, modifié d'après les données de la fig. 26, sur le poids des larves de têtes-de-boule.

minimale (DSM) satisfaisante (v. le § 7.2.4). Cette approche prudente serait satisfaisante.

On devrait signaler l'anomalie, que l'on ait réussi ou non à trouver une estimation ponctuelle. On devrait examiner les modes opératoires de l'essai pour trouver une cause, comme dans l'exemple 3.

Il est rare que l'on puisse imputer à un facteur biologique des anomalies reliées à la relation dose-effet et les actes d'agression en sont un exemple. Dans une série de 90 essais de criblage de toxicité létale effectués sur un effluent industriel, on a, dans certains d'entre eux, assisté à un comportement extrême d'agression chez les truites, après leur admission dans les enceintes expérimentales. Deux essais ont donné des résultats particulièrement étranges. Sur les 20 poissons par traitement, 9 sont morts chez le témoin et 5 dans la concentration minimale, apparemment par suite de combats. Aux deux concentrations intermédiaires, dans lesquelles les poissons ont semblé être tranquillisés par l'effluent, on n'a relevé aucune mortalité. Dans l'effluent non dilué, la toxicité a joué, et 16 poissons sont morts (Sprague, 1995). L'effet principal observé chez le témoin a été que, à l'évidence, un facteur étranger agissait. La relation déviante de la mortalité en fonction de la concentration, matérialisée par une courbe en U, pouvait s'expliquer mais ne pouvait pas être analysée par des moyens classiques.

(5) Courbe représentant un effet invariable. — La fig. 29 révèle un léger effet apparent à de nombreuses concentrations, mais ne montre aucune augmentation de l'effet en raison de la concentration. Manifestement, il y a une anomalie. Les résultats du côté droit pourraient être on ne pas être significativement inférieurs à ceux du témoin, mais l'expérimentateur devrait se méfier de l'invariabilité. Il ne devrait pas tenter d'estimer un paramètre de toxicité pour des données aussi extrêmes que celles de la figure.

On devrait rechercher la cause du phénomène dans le mode opératoire ou dans les facteurs biologiques. Voici certaines possibilités :

a) Les performances du témoin pourraient être exceptionnellement bonnes. On devrait les comparer à celles qui auront été constatées chez les témoins au laboratoire. Si les performances restent bonnes, les résultats de l'essai montrent simplement que l'effluent étudié n'est toxique à aucune concentration. (Il est peu probable que cela représente chez le témoin des performances qui ont été exceptionnellement mauvaises. Si tel était le cas, cela signifierait que la plupart ou toutes les

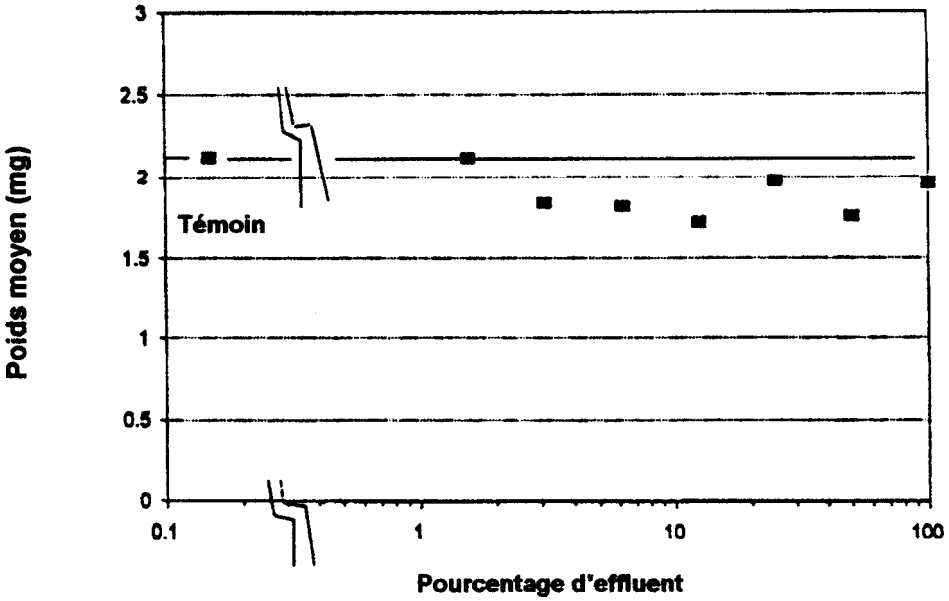


Figure 29. — Effet apparemment petit, mais variant à peine en fonction de la concentration. Cet exemple provient d'un laboratoire canadien ayant appliqué l'essai de mesure du poids de l'athérine, un poisson de mer.

concentrations expérimentales ont causé un effet, mais sans obéir à une relation concentration-effet.)

b) L'eau employée pour les dilutions pourrait ne pas avoir la qualité appropriée. Si, pour le témoin, on a utilisé un type d'eau (disons l'eau d'élevage) et si, pour les concentrations expérimentales, on a utilisé une autre eau, cela pourrait expliquer logiquement l'aplatissement de la distribution. Cette situation ne devrait pas se produire, puisque les modes opératoires préconisés par Environnement Canada exigent que l'eau de dilution et l'eau témoin soient la même eau. Dans l'exemple de fig. 29, l'explication pourrait se trouver dans quelque effet de la saumure ou des sels marins utilisés pour ajuster la salinité des concentrations expérimentales.

c) Il pourrait y avoir des effets pathogènes. Cela est peu probable, mais c'est possible dans des essais de toxicité chronique, particulièrement chez le poisson. Dans la matière à l'étude, il pourrait se trouver des pathogènes ayant agi sur les organismes en expérience, bien que la matière même n'ait pas été toxique. Si cela arrivait, les résultats seraient probablement plus erratiques que ceux que montre la fig. 27. Si la présence de pathogènes semble probable et si on voulait enquêter sur le phénomène, on pourrait effectuer des essais en parallèle, avec un essai englobant un traitement de la matière aux U. V. ou aux antibiotiques.

Si la courbe conservait cette allure dans un programme d'essais, il pourrait être souhaitable de mener une enquête par analyses chimiques ou par des techniques d'identification des agents toxiques.

(6) Relation inverse entre la concentration et l'effet. — À première vue, la fig. 30 pourrait représenter une relation ordinaire appropriée. À y regarder de plus près, on constate que les performances des cultures d'algues s'améliorent en raison de la concentration. La conclusion est simple : l'effluent n'est pas toxique pour les algues, mais il leur fournit des nutriments favorables à leur croissance et à leur reproduction. Une telle relation est le plus probable avec les végétaux, mais on pourrait aussi la constater chez d'autres organismes. (On devrait placer dans une perspective plus large la preuve de la présence de nutriments dans la matière à l'étude, pour ce qui concerne

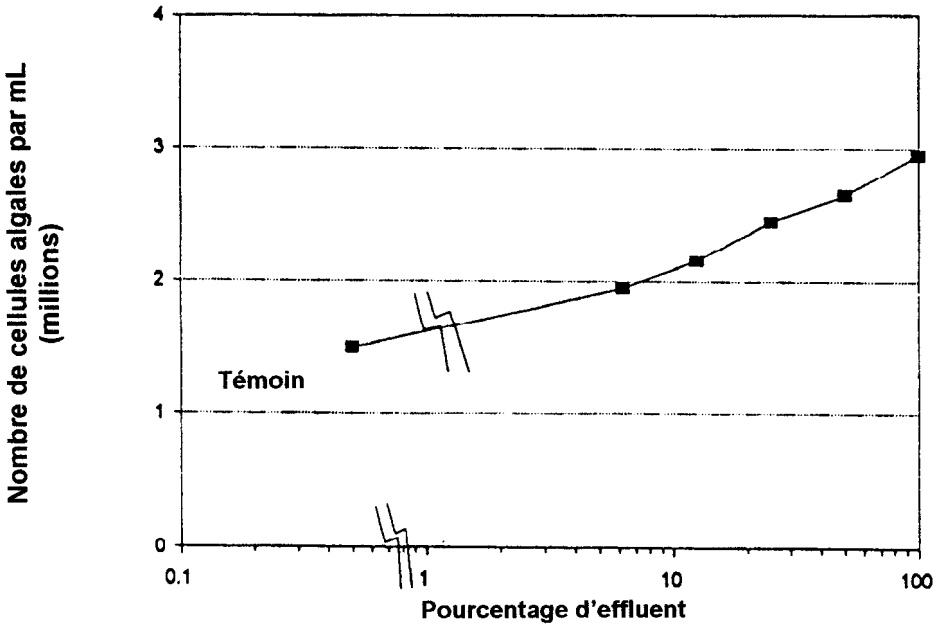


Figure 30. — Exemple de performances améliorées en fonction de la concentration. Cet exemple concerne le nombre de cellules algales. Il provient de l'USEPA (2000a).

l'enrichissement du milieu récepteur).

Une autre explication, peu probable mais possible, serait que la matière n'était pas toxique, mais que l'eau témoin ou l'eau de dilution l'était. Si, pour la dilution, on avait employé de l'eau du milieu récepteur, elle semblerait déjà toxique. Si cela était une explication possible et que la toxicité « absolue » de l'effluent ou d'une autre matière à l'étude doive être déterminée, on devrait utiliser une eau de dilution étalon, que l'on sait être inoffensive pour les organismes.

(7) Effets puissants à toutes les concentrations. — L'exemple de la fig. 31 montre des effets majeurs sur le nombre de cellules algales à toutes les concentrations expérimentales. La relation concentration-effet est presque horizontale. Ces résultats, non hypothétiques, sont bien réels. Manifestement, on aurait dû éprouver un intervalle plus grand de concentrations comme il en a été question aux alinéas 1 et 2. Les cinq concentrations utilisées ne couvrent qu'un ordre de grandeur. Si elles avaient été étalées davantage, les résultats pourraient avoir été moins énigmatiques. Avec les données actuelles, on ne peut pas estimer une CI 25 fiable, pas plus que l'on peut déterminer une approche fiable pour la CSEO et la CEMO.

Il est difficile de juger si les effets se seraient étendus aux concentrations plus fortes et plus faibles, vu la gamme étroite de concentrations montrée dans la fig. 31. Pour expliquer la distribution invariable, on pourrait aller jusqu'à invoquer un équilibre, à l'intérieur de la matière à l'étude, entre les constituants toxiques et d'autres qui stimuleraient la croissance des algues. Peut-être se trouverait-il une explication chimique à la quantité de forme active ou de constituant actif qui était libre d'agir aux diverses concentrations.

10.5 Interactions du mode opératoire sur les résultats

Repères

- *Le mode opératoire choisi pourrait influer sur l'analyse et les résultats.*
- *Dans un essai de croissance, par exemple, certaines mortalités survenant dans l'enceinte exposée à une forte concentration pourraient faire*

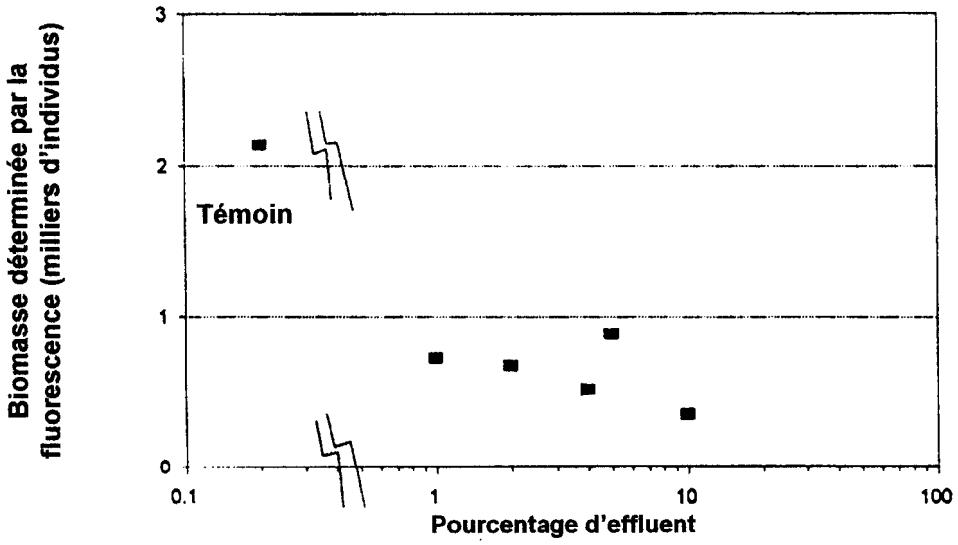


Figure 31. — Résultats d'un essai ne montrant que de grands effets. L'exemple provient d'un laboratoire canadien et concerne la biomasse de l'algue *Pseudokirchneriella subcapitata* [auparavant *Selenastrum capricornutum*].

en sorte que les survivants, disposant de plus de nourriture, présenteraient une croissance compensatoire, ce qui compromettrait les conclusions de l'analyse et occulterait les effets sublétaux.

- *De même, une mortalité partielle dans l'enceinte pourrait faire en sorte que les survivants seraient exposés à plus de toxique, ce qui augmenterait les effets sublétaux.*
- *La meilleure parade contre de telles influences consiste à utiliser des modes opératoires éprouvés. La modification du régime alimentaire et des taux de renouvellement favorables de la solution pourrait éviter les problèmes mentionnés.*

Des questions peuvent être soulevées concernant l'influence des méthodes d'essai elles-mêmes sur l'analyse statistique et l'estimation du paramètre de toxicité. L'interaction peut aboutir à des résultats qui ne conviennent pas à l'analyse statistique ou qui sont difficiles à interpréter. Le sujet est rarement pris en considération, mais, parfois, il pourrait être important. Nous en donnons un exemple ; il pourrait y en avoir des analogues pour d'autres essais et d'autres effets.

Le nombre d'organismes par récipient pourrait facilement influer sur l'analyse et les résultats dans les essais de toxicité sublétale. Si, dans chaque récipient, il se trouvait plusieurs organismes et que, dans certains récipients mais non dans d'autres, ces organismes mouraient, cela pourrait conduire à des expositions déséquilibrées, qui influeraient sur les effets. Le traitement inégal des groupes pourrait aller à l'encontre des exigences en matière d'analyse statistique.

- Si, dans un récipient, la plupart des organismes mouraient, les survivants profiteraient-ils du surcroît de nourriture à leur disposition ? L'évaluation de la croissance pourrait-elle être biaisée ? Cela est certes possible si la technique d'alimentation fournit plus de nourriture à chaque survivant. On connaît des exemples évidents d'absorption alimentaire compensatoire (accrue) chez le poisson, qui peut neutraliser les effets nuisibles d'un toxique sur la croissance (Warren, 1971). Un rationnement calculé d'après le nombre d'organismes ou leur biomasse pourrait corriger ce problème.

- Le choix de la ration pourrait facilement influer sur le résultat obtenu, vu le phénomène précédemment mentionné de croissance compensatoire.

L'expérimentateur pourrait choisir une ration relativement importante, dans l'espoir de montrer les différences maximales existant entre les concentrations expérimentales, mais une ration trop généreuse pourrait gommer les différences, en raison de l'alimentation compensatoire.

- Si la plupart des organismes dans un récipient mouraient, les survivants seraient-ils exposés à une dose plus forte ? Vraisemblablement, les morts cesseraient d'assimiler du toxique, de sorte qu'ils n'abaisseraient pas les concentrations ambiantes. L'exposition des survivants serait plus grande que s'il n'y avait pas eu de mortalité. L'effet pourrait être ou ne pas être négligeable.

La principale parade contre ces anomalies consiste à employer un mode opératoire éprouvé. Les méthodes normalisées telles que celles d'Environnement Canada sont désormais largement accessibles ; les méthodes sont efficaces, ayant généralement été raffinées par des groupes chevronnés. Les rations auront été choisies pour réduire au minimum les anomalies. Pour maintenir les concentrations voulues dans les enceintes expérimentales, toute influence due à des mortalités partielles dans des groupes d'organismes serait neutralisée par les volumes importants de solution d'essai utilisée pour la biomasse présente.

Références

Abbott, W.S., 1925. A method of computing the effectiveness of an insecticide. *J. Econ. Ent., 18:*265–267.

Alderdice, D.F. et J.R. Brett, 1957. Some effects of kraft mill effluent on young Pacific salmon. *J. Fish. Res. Board Can., 14:*783–795.

Andersen, H. 1994. *Statistical methods for evaluation of the toxicity of waste water.* Thèse de maîtrise en sciences, Section de modélisation mathématique, Université technique du Danemark à Lyngby [en danois].

Andersen, J.S., H. Holst, H. Spliid, H. Andersen, A. Baun et N. Nyholm, 1998. Continuous ecotoxicological data evaluated relative to a control response. *J. Agric. Biol. and Environ. Statistics, 3:*405–420.

Andersen, J.S., J.J.M. Bedaux, S.A.L.M. Kooijman et H. Holst, 2000. The influence of design characteristics on statistical inference in non-linear estimation; a simulation study. *J. Agric. Biol. and Environ. Statistics, 5:*28–48.

Anonyme, 1994. How to measure no-effect? *SETAC News,* nov. 1994 : p. 19. [Society Environ. Toxicol. and Chemistry]

APHA, AWWA et WEF, 1992 [American Public Health Association, American Water Works Association et Water Environment Federation]. *Standard methods for the examination of water and wastewater.* 18^{e} éd. APHA, Washington.

Ashton, W.D., 1972. *The logit transformation with special reference to its uses in bioassay.* Griffin's Statistical Monographs & Courses, n^{o} 332. Hafner Pub. Co., New York, 88 p.

Atkinson, G.F., 1999. Assessment of available computer programs. Attachment T, 2 p., dans : *Minutes/Proceedings of the Statistics Workshop for Toxicological Testing,* Pacific Environmental Science Centre (PESC), North Vancouver B.C., September 15–17th, 1999. Environnement Canada, Centre des sciences environnementales du Pacifique, North Vancouver, C.-B.

Bailer, A.J et J.T. Oris, 1993. Modeling reproductive toxicity in *Ceriodaphnia* tests. *Environ. Toxicol. Chem. 12:*787–791.

————. 1994. Assessing toxicity of pollutants in aquatic systems. p. 28–40, dans : *Case studies in biometry.* N. Lange, L. Ryan, L. Billard, D. Brillinger, L. Conquest et J. Greenhouse (dir.). John Wiley & Sons, Inc., New York.

————. 1997. Estimating inhibition concentrations for different response scales using generalized linear models. *Environ. Toxicol. Chem., 16:*1554–1559.

————. 1999. What is an NOEC? Non-monotonic concentration-response patterns want to know. *SETAC News, March 1999:*22–24.

Bailer, A.J., M.R. Hughes, D.L. Denton et J.T. Oris, 2000a. An empirical comparison of effective concentration estimators for evaluating aquatic toxicity test responses. *Environ. Toxicol. Chem., 19:*141–150.

Bailer, A.J., R.T. Elmore, B.J. Shumate et J.T. Oris, 2000b. Simulation study of characteristics of statistical estimators of inhibition concentration. *Environ. Toxicol. Chem., 19:*3068–3073.

Baird, R.B., R. Berger et J. Gully, 1995. Improvements in point estimation methods and application to controlling aquatic toxicity test reliability. p. 103–130, dans : *Whole effluent toxicity testing as evaluation of methods and prediction of receiving system impacts.* D.R. Grothe, K.L. Dickson et K.K. Reed-Judkins (dir.), SETAC Press, Pensacola, Floride.

Barnett, V. et F. Lewis, 1994. *Outliers in statistical data.* 3^{e} éd. Wiley, New York.

Bartlett, M.S., 1937. Some examples of statistical methods of research in agriculture and applied biology. *J. Roy. Stat. Soc. Suppl., 4:*137–170.

Bates, D.M. et D.G. Watts, 1988. *Nonlinear regression analysis and its applications.* John Wiley & Sons, New York, 365 p.

Beyers, D.W., T.J. Keefe et C.A. Carlson, 1994. Toxicity of Carbaryl and Malathion to two federally endangered fishes, as estimated by regression and ANOVA. *Environ. Toxicol. Chem., 13:*101–107.

Billington, J.W., G.-L Huang, F. Szeto, W.Y. Shiu et D. MacKay, 1988. Preparation of aqueous solutions of sparingly soluble organic substances: I. Single component systems. *Environ. Toxicol. Chem., 7:*117–124.

Bliss, C.I., 1937. The calculation of the time-mortality curve. *Ann. Appl. Biol., 24:*815–852.

Bliss, C.I. et McK. Cattell, 1943. Biological assay. *Ann. Rev. Physiol., 5:*479–539.

Borgmann, U., 1994. Chronic toxicity of ammonia to the amphipod *Hyalella azteca*; importance of ammonium ion and water hardness. *Environ. Pollut., 86:*329–335.

Brain, P. et R. Cousens, 1989. An equation to describe dose responses where there is stimulation of growth at low doses. *Weed Res., 29:*93–96.

Broderius, S.J., 1991. Modeling the joint toxicity of xenobiotics to aquatic organisms: basic concepts and approaches. p. 107–127, dans : *Aquatic toxicology and risk assessment: fourteenth volume.* ASTM STP 1124, M.A. Mayes et M.G. Barron (dir.), Amer. Soc. Testing and Materials, Philadelphie.

Bruce, R.D. et D.J. Versteeg, 1992. A statistical procedure for modeling continuous toxicity data. *Environ. Toxicol. Chem., 11:*1485–1494.

Buikema, A.L., Jr., B.R. Niederlehner et J. Cairns, Jr., 1982. Biological monitoring. Part IV -- Toxicity testing. *Water Res., 16:*239–262.

Burchfield, R.W., 1996. *The new Fowler's modern English usage*. 3e éd. Clarendon Press, Oxford.

Calabrese, E.J. et L.A. Baldwin, 1997. The dose determines the stimulation (and poison). Development of a chemical hormesis database. *Int. J. Toxicol., 16:*545–559.

Calamari, D., R. Marchetti et G. Vailati, 1980. Influence of water hardness on cadmium toxicity to *Salmo gairdneri* Rich. *Water Research, 14:*1421–1426.

Carter, E.M. et J.J. Hubert, 1984. A growth-curve model approach to multivariate quantal bioassay. *Biometrics, 40:*699–700.

Caux, P.Y. et D.R.J. Moore, 1997. A spreadsheet program for estimating low toxic effects. *Environ. Toxicol. Chem., 16:*802–806.

CCREM [Conseil canadien des ministres des ressources et de l'environnement], 1987. *Recommandations canadiennes pour la qualité des eaux.* CCMRE, Groupe de travail sur les recommandations pour la qualité des eaux. Environnement Canada, Ottawa.

CETIS, 2001. Comprehensive Environmental Toxicity Information System. Tidepool Scientific Software, McKinleyville, Calif. 95521 [Programme sur disquette et guide imprimé de l'utilisateur.]

Chapman, P.M., 1996. *Alternatives to the NOEC based on regression analysis.* Document de travail, annexe 7, OECD Workshop on Statistical Analysis of Aquatic Ecotoxicity Data, Brunswick, Allemagne, du 15 au 17 oct. 1996, 5 p. Organisation de coopération et de développement économiques, Paris. [Annexe A dans Moore, 1996.]

Chapman, P.F., M. Crane, J. Wiles, F. Noppert et E. McIndoe, 1996a. Asking the right questions: ecotoxicology and statistics. SETAC-Europe, Bruxelles [Society of Environmental Toxicology and Chemistry]. Compte rendu d'un atelier tenu au Royal Holloway University of London, Egham, Surrey, R.-U., 26-27 avril 1995.

Chapman, P.M., R.S. Caldwell et P.F. Chapman, 1996b. A warning: NOECs are inappropriate for regulatory use. *Environ. Toxicol. Chem., 15:*77–79.

Christensen, E.R., 1984. Dose-response functions in aquatic toxicity testing and the Weibull model. *Wat. Res., 18:* 213–221.

Christensen, E.R. et N. Nyholm, 1984. Ecological assays with algae: Weibull dose-response curves. *Env. Sci. Technol., 19:*713–718

Cochran, W.G. G.M. Cox, 1957. *Experimental designs.* 2e éd. Wiley, New York, 611 p.

Cohen, J., 1964. Psychological time. *Scientific Amer., 211, No 5:*117–118.

Collett, D., 1991. *Modelling binary data*. Chapman & Hall, Londres. 369 p.

Crane, M. et E. Godolphin, 2000. Statistical analysis of effluent bioassays. Environment Agency, Bristol, U.K. Research and Development Tech. Rept E19.

Crane, M. et M.C. Newman, 2000. What level of effect is a no observed effect? *Environ. Toxicol. Chem., 19:*516–519.

Crane, M., M.C. Newman, P.F. Chapman et J. Fenlon, 2002. *Risk assessment with time to event models*. Lewis Publishers/CRC Press, Boca Raton, Floride, 302 p.

Critchlow, D.E. et M.A. Fligner, 1991. On distribution-free multiple comparisons in the one-way analysis of variance. *Comm. Stat. Theory Methods, 20:*127–139.

D'Agostino, R.B., 1986. Tests for the normal distribution. p. 367–420, dans : *Goodness-of-fit techniques.* R.B. D'Agostino et M.A. Stephens (dir.), Marcel Dekker Inc., New York.

Damico, J.A. et D.A. Wolfe, 1987. Extended tables of the exact distribution of a rank statistic for treatment versus control multiple comparisons in one-way layout designs. *Comm. Stat. Theory Methods, 18:*3327–3353.

Davis, J.M. et D.J. Svendsgaard, 1990. U-shaped does-response curves: their occurrence and implications for risk assessment. *J. Toxicol. & Environ. Health, 30:*71–83.

Davis, R.B., A.J. Bailer et J.T. Oris, 1998. Effects of organism allocation on toxicity test results. *Environ. Toxicol. Chem., 17:*928–931.

deBruijn, H.H.M. et M. Hof, 1997. How to measure no effect. Part IV: How acceptable is the ECx from an environmental policy point of view? *Environmetrics, 8:* 263–267.

Dixon, W.J. et F.J. Massey Jr., 1983. *Introduction to statistical analysis.* 4e éd. McGraw-Hill, New York.

Dixon, P.M. et M.C. Newman, 1991. Analyzing toxicity data using statistical models for time-to-death: an introduction. p. 207–242, dans : *Metal ecotoxicology, concepts and applications*. M.C. Newman et A.W. McIntosh (dir.). Lewis Publishers, Inc., Chelsea, Mich., 399 p.

Dobson, A.J., 2002. *An introduction to generalized linear models.* 2e éd. Chapman & Hall/CRC, Boca Raton, Floride, et Londres, 240 p.

Doe, K.G., 1994. Comments on the minutes of the Toxicological Statistics Advisory Group Meeting in Quebec City. Note à J.A. Miller, Direction du développement technologique, 28 juillet 1994. [K.G. Doe, chef, Section de toxicologie, Environnement Canada, Dartmouth, N.-É.]

Douglas, M.T., D.O. Chanter, I.B. Pell et G.M. Burney, 1986. A proposal for the reduction of animal numbers required for the acute toxicity to fish test (LC_{50} determination). *Aquat. Toxicol., 8:*243–249.

Draper, N.R. et J.A. John, 1981. Influential observations and outliers in regression. *Technometrics, 23:*21–26.

Draper, N.R. et H. Smith, 1981. *Applied regression analysis.* 2e éd. Wiley, New York, 709 p.

Dunnett, C.W., 1955. A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Stat. Assoc., 50:*1096–1121.

———. 1964. New tables for multiple comparisons with a control. *Biometrics, 20:*482–491.

Dunnett, C.W. et A.C. Tamhane, 1998. New multiple test procedures for dose finding. *J. Biopharmaceut. Stat., 8:* 353–366.

Du Nouy, L., 1936. *Biological time.* Methuen, Londres. 180 p.

EC [Environnement Canada], 1990a. Méthode d'essai biologique : Essai de létalité aiguë sur la truite arc-en-ciel. Série de la protection de l'environnement. Ottawa, publication SPE 1/RM/9 (modifié en 1996).

———. 1990b. Méthode d'essai biologique : Essai de létalité aiguë sur l'épinoche à trois épines *Gasterosteus aculeatus*. Série de la protection de l'environnement, Ottawa, publication SPE 1/RM/10 (modifié en 2002).

———. 1990c. Méthode d'essai biologique : Essai de létalité aiguë sur *Daphnia* sp. Série de la protection de l'environnement, Ottawa, publication SPE 1/RM/11 (modifié en 1996).

———. 1990d. Document d'orientation sur le contrôle de la précision des essais de toxicité au moyen de produits toxiques de référence. Série de la protection de l'environnement, Ottawa, publication SPE 1/RM/12.

———. 1992a. Méthode d'essai biologique : Essai de reproduction et de survie sur le cladocère *Ceriodaphnia dubia*. Série de la protection de l'environnement, Ottawa, publication SPE 1/RM/21 (modifié en 1997).

———. 1992b. Méthode d'essai biologique : Essai de croissance et de survie sur des larves de tête-de-boule. Série de la protection de l'environnement, Ottawa, publication SPE 1/RM/22 (modifié en 1997).

———. 1992c. Méthode d'essai biologique : Essai de toxicité sur la bactérie luminescente *Photobacterium phosphoreum*. Série de la protection de l'environnement, Ottawa, publication SPE 1/RM/24.

———. 1992d. Méthode d'essai biologique : Essai d'inhibition de la croissance de l'algue d'eau douce *Selenastrum capricornutum*. Série de la protection de l'environnement, Ottawa, publication SPE 1/RM/25 (modifié en 1997).

———. 1992e. Méthode d'essai biologique : Essai de toxicité aiguë de sédiments chez les amphipodes marins ou estuariens. Série de la protection de l'environnement, Ottawa, publication SPE 1/RM/26 (modifié en 1998).

———. 1992f. Méthode d'essai biologique : Méthode d'essai biologique : essai sur la fécondation chez les échinides (oursins verts et oursins plats). Série de la protection de l'environnement, Ottawa, publication SPE 1/RM/27 (modifié en 1997).

———. 1992g. Fertilization assay with echinoids: interlaboratory evaluation of test options. EC, Conservation et protection, Direction du développement technologique, Ottawa. Inédit, 45 p. + ann.

———. 1994. Document d'orientation sur le prélèvement et la préparation de sédiments en vue de leur caractérisation physico-chimique et d'essais biologiques, Section de l'élaboration et de l'application des méthodes, Ottawa, publication SPE 1/RM/29.

———. 1997a. Méthode d'essai biologique : essai de survie et de croissance des larves dulcicoles de chironomes *(Chironomus tentans* ou *Chironomus riparius*) dans les sédiments, Section de l'élaboration et de l'application des méthodes, Ottawa, publication SPE 1/RM/32.

———. 1997b. Méthode d'essai biologique : essai de survie et de croissance de l'amphipode dulcicole *Hyalella azteca* dans les sédiments, Section de l'élaboration et de l'application des méthodes, Ottawa, publication SPE 1/RM/33.

———. 1998a. Méthode d'essai biologique : essais toxicologiques sur des salmonidés (truite arc-en-ciel) aux premiers stades de leur cycle biologique. 2e éd. Section de l'élaboration et de l'application des méthodes, Ottawa, publication SPE 1/RM/28.

———. 1998b. Méthode d'essai biologique : méthode de référence pour la détermination de la létalité aiguë d'un sédiment pour des amphipodes marins ou estuariens, Section de l'élaboration et de l'application des méthodes, Ottawa, publication SPE 1/RM/35.

———. 1999a. Guide des essais écotoxicologiques employant une seule espèce et de l'interprétation de leurs résultats, Section de l'élaboration et de l'application des méthodes, Ottawa, publication SPE 1/RM/34.

———. 1999b. Méthode d'essai biologique : essai de mesure de l'inhibition de la croissance de la plante macroscopique dulcicole *Lemna minor,* Section de l'élaboration et de l'application des méthodes, Ottawa, publication SPE 1/RM/37.

———. 2000a. Méthode d'essai biologique : Méthode de référence pour la détermination de la létalité aiguë d'effluents chez la truite arc-en-ciel. 2e éd. Section de l'élaboration et de l'application des méthodes, Ottawa, publication SPE 1/RM/13.

———. 2000b. Méthode d'essai biologique : Méthode de référence pour la détermination de la létalité aiguë d'effluents chez *Daphnia magna*. 2e éd. Section de l'élaboration et de l'application des méthodes, Ottawa, publication SPE 1/RM/14.

———. 2001a. Méthode d'essai biologique : Essai de survie et de croissance des vers polychètes spionides *(Polydora cornuta)* dans les sédiments, Section de l'élaboration et de l'application des méthodes, Ottawa, publication SPE 1/RM/41.

———. 2001b. Revised procedures for adjusting salinity of effluent samples for marine sublethal toxicity testing conducted under Environmental Effects Monitoring (EEM) programs, Section de l'élaboration et de l'application des méthodes, Ottawa. Non numéroté, octobre 2001, 9 p.

———. 2002a. Méthode d'essai biologique : Méthode de référence servant à déterminer la toxicité des sédiments à l'aide d'une bactérie luminescente dans en un essai en phase solide, Section de l'élaboration et de l'application des méthodes, Ottawa, publication SPE 1/RM/42.

———. 2002b. Guide pour l'étude du suivi des effets sur l'environnement aquatique par les mines de métaux. Service de la conservation de l'environnement, Ottawa.

———. 2004a. Méthode d'essai biologique : Essais pour déterminer la toxicité de sols contaminés pour les vers de terre *Eisenia andrei, Eisenia fetida* ou *Lumbricus terrestris,* Section de l'élaboration et de l'application des méthodes, Ottawa, publication SPE 1/RM/43.

———. 2004b. Méthode d'essai biologique : Essais de mesure de la levée et de la croissance de plantes terrestres exposées à des contaminants dans le sol, Section de l'élaboration et de l'application des méthodes, Ottawa, publication SPE 1/RM/45.

———. 2007. Méthode d'essai biologique : Essai de mesure de la survie et de la reproduction de collemboles exposés à des contaminants dans le sol. Section de l'élaboration et de l'application des méthodes, Ottawa, publication SPE 1/RM/47.

Edwards, D. et J.J. Berry, 1987. The efficiency of simulation-based multiple comparisons. *Biometrics, 43:* 913–926.

Efron, B., 1982. The jackknife, the bootstrap and other resampling plans. Soc. Indust. Appld. Math, Philadelphie., CMMS 38.

Finney, D.J., 1971. *Probit analysis.* 3e éd. Cambridge University Press, Londres. 333 p.

———. 1978. *Statistical method in biological assay*. 3e éd. Charles Griffin & Co. Ltd, Londres. 508 p.

Finney, D.J., R. Latscha, B.M. Bennett et P. Hsu, 1963. *Tables for testing significance in a 2 × 2 contingency table.* Published for the Biometrika Trustees, at the University Press, Cambridge, 102 p.

Fleiss, J.L., 1981. *Statistical methods for rates and proportions.* 2nd Edition, John Wiley & Sons, Toronto. 321 p.

Fligner, M.A. et D.A. Wolfe, 1982. Distribution-free tests for comparing several treatments with a control. *Stat. Neer., 36:*119–127.

Fry, F.E.J. 1947. Effects of the environment on animal activity. Univ. Toronto Studies, Biol. Series no. 55, *Publ. Ont. Fish. Res. Lab., 68:*1–62

Gad, S.C., 1999. *Statistics and experimental design for toxicologists.* CRC Press, Boca Raton, Floride. 437 p.

Gaddum, J.H., 1953. Bioassays and mathematics. *Pharmacol. Rev., 5:*87–134.

Gelber, R.D., P.T. Lavin, C.R. Mehta et D.A. Schoenfeld, 1985. Statistical analysis. p. 110–123, dans : *Fundamentals of aquatic toxicology. Methods and applications.*, G.M. Rand et S.R. Petrocelli (dir.), Hemisphere Publishing Corporation, Washington, D.C.

Grothe, D.R., K.L. Dickson et D.K. Reed-Judkins, 1996. Whole effluent toxicity testing: An evaluation of methods and prediction of receiving system impacts. SETAC Press (Soc. Environmental Toxicol. and Chemistry] Pensacola, Floride. 346 p.

Grubbs, F.E., 1969. Procedures for detecting outlying observations in samples. *Technometrics, 11:*1–21.

Hamilton, M.A., 1979. Robust estimates of the ED50. *J. Amer. Stat. Assoc., 74:*344–354.

———. 1980. Inference about the ED50 using the trimmed Spearman-Kärber procedure -- a Monte Carlo investigation. *Commun. Statist. Simula. Computa. B. 9(3):*235–254.

———. 1986. Statistical analysis of the cladoceran reproductivity test. *Environ. Toxicol. Chem., 5:*205–212.

Hamilton, M.A., R.C. Russo et R.V. Thurston, 1977. Trimmed Spearman-Kärber method for estimating median lethal concentrations in toxicity bioassays. *Environ. Sci. Technol.11:*714–719. Errata, même périodique. *12:*417.

Härdle, W., 1991. *Smoothing techniques with implementation in S.* Springer-Verlag, New York..

Hastie, T. et R. Tibshirani, 1990. *Generalized additive models.* Chapman and Hall, Londres.

Hayter, A.J. et G. Stone, 1991. Distribution-free multiple comparisons for monotonically ordered treatment effects. *Austral. J. Stat., 33:*335–346.

Heming, T.A., S. Arvind et K. Yogesh, 1989. Time-toxicity relationships in fish exposed to the organochlorine pesticide methoxychlor. *Environ. Toxicol. Chem., 8:* 923–932.

Hewlett, P.S. et R.L. Plackett, 1979. *The interpretation of quantal responses in biology.* University Park Press, Baltimore, Maryland. 82 p.

Hochberg, Y. et A.C. Tamhane, 1987. *Multiple comparison procedures.* J. Wiley and Sons, New York.

Hodson, P.V., C.W. Ross, A.J. Niimi et D.J. Spry, 1977. Statistical considerations in planning aquatic bioassays. p. 15–31, dans : *Proc. 3rd Aquatic Toxicity Workshop*, Halifax, 2–3 nov. 1976. Environnement Canada, Service de la protection de l'environnement, rapp. techn. EPS-5-AR-77-1, Halifax.

Hoekstra, J.A., 1989. Estimation of the ED50; lettre au directeur de la publication. *Biometrics, 45:*337–338.

Hoekstra, J.A. et P.H. Van Ewijk, 1993. Alternatives for the no-observed-effect level. *Environ. Toxicol. Chem., 12:*187–194.

Hollander, M. et D.A. Wolfe, 1999. *Nonparametric statistical methods.* J. Wiley and Sons, New York, 787 p.

Hong, W-H., P.G. Meier et R.A Deininger, 1988. Determination of dose-time response relationships from long-term acute toxicity test data. *Environ. Toxicol. Chem., 7:*221–226.

Horness, B.H., D.P. Lomax, L.L. Johnson, M.S. Myers, S.M. Pierce et T.K. Collier, 1998. Sediment quality thresholds: estimates from hockey stick regression of liver lesion prevalence in English sole *(Pleuronectes vetulus)*. *Environ. Toxicol. Chem., 17:*872–882.

Hosmer, D.W. et S. Lemeshow, 2000. *Applied logistic regression.* 2e éd. Wiley-Interscience, New York.

Hubert, J.J., 1984. *Bioassay.* 2e éd. Kendall/Hunt Pub. Co., Dubuque, Iowa, 180 p.

———. 1992. *Bioassay.* 3e éd. Kendall/Hunt Pub. Co., Dubuque, Iowa, 198 p.

———. 1987. *PROBIT2: A microcomputer program for probit analysis.* Département de mathématiques et de statistique, U. de Guelph, Guelph ON N1G 2W1.

Hurlbert, S.H., 1984. Pseudoreplication and the design of ecological field experiments. *Ecol. Monog., 54:*187–2112.

ISO [Organisation internationale de normalisation], 1999. Water quality -- Guidlines for alagal growth inhibition tests with poorly soluble materials, volatile compounds, metals and waste water. ISO, Genève. ISO 14442, 14 p.

Jackman, P. et K. Doe, 2003. Evaluation of CETIS statistical software. Rapport non numéroté, Environnement Canada, Centre des sciences de l'environnement, Moncton. 46 p.

Jennrich, R.I. et R.H. Moore, 1975. Maximum likelihood estimation by means of nonlinear least squares. Proc. *Statistical Computing Section, Amer. Statistical Assoc., 70:*57–65.

Jensen, A.L., 1972. Standard error of LC_{50} and sample size in fish bioassays. *Water Res., 6:*85–89.

Jonckheere, A.R., 1954. A distribution free k-sample test against ordered alternatives. *Biometrika, 41:*133–145.

Kappenman, R.F., 1987. Nonparametric estimation of dose-response curves with application to ED50 estimation. *J. Stat. Com. Sim., 28:*1–13.

Kerr, D.R. et J.P. Meador, 1996. Modeling dose response using generalized linear models. *Environ. Toxicol. Chem., 15:*395–401.

Kooijman, S.A.L.M., 1996. An alternative for NOEC exists, but the standard model has to be abandoned first. *Oikos 75:* 310–316.

Kooijman, S.A.L.M. et J.J.M Bedaux, 1996. The analysis of aquatic toxicity data. VU Univ. Press, Vrije Universiteit, Amsterdam. 149 p. [Comprend la disquette du logiciel DEBtox.]

Kooijman, S.A.L.M., A.O. Hanstveit et N. Nyholm, 1996. No-effect concentration in algal growth inhibition tests. *Water Res., 30:*1625–1632.

Koper, N., 1999. Nonlinear regression lecture for Vancouver workshop. Pièce Ra, 12 p., dans : *Minutes/Proceedings of the Statistics Workshop for Toxicological Testing,* Centre des sciences environnementales du Pacifique (CSEP), North Vancouver C.-B., 15–17 sept. 1999. Environnement Canada, Centre des sciences environnementales du Pacifique, North Vancouver.

Kruskal, W.H. et W.A. Wallis, 1952. Use of ranks in one-criterion analysis of variance. *J. Amer. Statist. Assoc., 47:*583–621.

Lanno, R.P., G.L. Stephenson et C.D. Wren, 1997. Applications of toxicity curves in assessing the toxicity of diazinon and pentachlorophenol to *Lumbricus terrestris* in natural soils. Soil Biology and Biochemistry 29: 689-692.

Lee, G., M.R. Ellersieck, F.L. Mayer, et G.F. Krause, 1995. Predicting chronic lethality of chemicals to fishes from acute toxicity test data: multifactor probit analysis. *Environ. Toxicol. Chem., 14:*345–349.

Levene, H., 1960. Robust tests for the equallity of variances. p. 278–292, dans : *Contributions to probability and statistics.* I. Olkin (dir.), Stanford Univ. Press, Palo Alto, Calif.

Litchfield, J.T., 1949. A method for rapid graphic solution of time-percent effect curves. *Pharmacol. Exp. Ther., 97:*399–408.

Litchfield, J.T. et F. Wilcoxon, 1949. A simplified method of evaluating dose-effect experiments. *J. Pharmacol. Experimental Therapeutics, 96:*99–113.

Lloyd, Richard, 1992. *Pollution and freshwater fish. Fishing News Books* (Blackwell Scientific Publications Ltd), Oxford. 176 p.

Mallows, C.L., 1973. Some comments on C_p. *Technometrics, 12:*621–625.

Manly, B.F.J., 2000. *Statistics for environmental science and management.* CRC Press, Boca Raton, Floride. 336 p.

Marcus, A.H. et A.P. Holtzman, 1988. A robust statistical method for estimating effects concentrations in short-term fathead minnow toxicity tests. Battelle Washington Environmental Program Office, Washington, D.C. Report for USEPA Office of Water, Contract n° 69-03-3534, 39 p.

McCullagh, P. et J.A. Nelder, 1989. *Generalized linear models.* Chapman & Hall/CRC, Boca Raton, Floride. 532 p.

———. 1994. *Generalized linear models.* 2nd Chapman & Hall/CRC, Boca Raton, Floride, et Londres. 511 p.

McLeese, D.W., 1956. Effects of temperature, salinity and oxygen on the survival of the American lobster. *J. Fish. Res. Bd Canada, 13:*494–502.

Millard, S.P. et N.K. Neerchal, 2000. *Environmental statistics with S-Plus.* CRC Press, Boca Raton, Floride. 848 p.

Miller, R.G., 1981. *Simultaneous statistical inference.* Springer-Verlag, New York. 299 p.

———. 1986. *Beyond ANOVA, basics of applied statistics.* John Wiley & Sons, New York. [Cité par Newman, 1995.]

Miller, R.G. et J.W. Halpern, 1980. Robust estimators for quantal bioassay. *Biometrika, 67:*103–110.

Miller, J., R.P. Scroggins et G.F. Atkinson, 1993. Toxicity endpoint determination statistics and computer programs. Compte rendu de la réunion du Groupe consultatif de la statistique à Québec, le 20 oct. 1993. Environnement Canada, Direction du développement technologique, Ottawa. 12 p. + annexes.

Moody, M. 2003. Research to assess potential improvements to Environment Canada's *Lemna minor* test method. Saskatchewan Research Council, Saskatoon, Sask., publication 11545-1C03. 69 p.

Moore D.R.J., 1996. *OECD workshop on statistical analysis of aquatic ecotoxicity data.* Rapport sommaire pour Environnement Canada. Non numéroté, 31 oct. 1996, The Cadmus Group, Ottawa, 10 p. + annexes.

Moore, D.R.J. et P.-Y. Caux, 1997. Estimating low toxic effects. *Environ. Toxicol. Chem., 16:*794–801.

Moore, T.F., S.P. Canton et M. Grimes, 2000. Investigating the incidence of type 1 errors for chronic whole effluent toxicity testing using *Ceriodaphnia dubia*. *Environ. Toxicol. Chem,. 19:*118–122.

Morissette, S., 2002. Le coût de l'incertitude en échantillonnage environnemental. Annexe C, dans : *Environnement Canada. Guide d'échantillonnage des sédiments du Saint-Laurent pour les projets de dragage et de génie maritime. Vol. 1: Directives de planification.* Environnement Canada, Direction de la protection de l'environnement, Région du Québec, Section innovation technologique et secteurs industriels. Rapport 106 p. [accessible sur http://www.slv2000.qc.ca.]

Müller, H.-G. et T. Schmitt, 1988. Kernel and probit estimates in quantal bioassay. *J. Amer. Stat. Assoc., 83:* 750–758.

Newman, M.C., 1995. *Quantitative methods in aquatic ecotoxicology.* Lewis Pub., Boca Raton, Floride. 426 p.

Newman, M.C. et M.S. Aplin, 1992. Enhancing toxicity data interpretation and prediction of ecological risk with survival time modeling: an illustration using sodium chloride toxicity to mosquitofish *Gambusia holbrooki*. *Aquatic Toxicol., 23:*85–96.

Noppert, F., N. van der Hoeven et A. Leopold (dir.), 1994. How to measure no effect. Towards a new measure of chronic toxicity in ecotoxicology. Compte rendu d'atelier, La Haye, 9 sept. 1994. Groupe de travail néerlandais sur la statistique et l'écotoxicologie [Copies à : BKH Consulting Engineers, P.O. box 5094, 2600 GB, Delft, The Netherlands, att. F. Noppert.]

Norberg-King, T.J., 1993. A linear interpolation method for sublethal toxicity: the Inhibition Concentration (ICp) approach (Version 2.0). USEPA, Duluth, Minn., Tech. Rept 03-93, National Effluent Toxicity Assessment Center, 25 p.

Nyholm, Niels, 2001. Laboratoire des sciences et de l'écologie, U. technique du Danemark à Lyngby. Observations sur une ébauche antérieure du présent document. Communication personnelle.

Nyholm, N., P.S. Sørensen, K.O. Kusk et E.R. Christensen, 1992. Statistical treatment of data from microbial toxicity tests. *Environ. Toxicol. Chem., 11:* 157–167.

O'Brien, R.G., 1979. A general ANOVA method for robust tests of additive models for variances. *J. Amer. Stat. Assoc., 74:* 877–880.

OECD [Organisation de coopération et de développement économiques (OCDE)], 1995. Guidance document for aquatic effects assessment. OCDE, Paris. OECD Environment Monographs No. 92, 116 p.

———. 1997. Report of the final ring-test of the *Daphnia magna* reproduction test. OCDE, Paris. OECD Environmental Health and Safety Publications, Series on Testing and Assessment No. 6.

———. 1998. Report of the OECD workshop on statistical analysis of aquatic toxicity data. OCDE, Paris. OECD Environmental Health and Safety Publications, Series on Testing and Assessment No. 10, 133 p.

———. 2004. Draft guidance document on the statistical analysis of ecotoxicity data. OCDE, Paris, Environmental Health and Safety Pub., Series on Testing and Assessment. 214 p. [accessible à www.oecd.org]

OMEE [Ministère de l'Environnement et de l'Énergie de l'Ontario (MEEO)], 1995. TOXSTATS. OMEE, Etobicoke, Ont. [Programmes permettant l'estimation de la CE 50 en format Windows.]

Pack, S., 1993. *A review of statistical data analysis and experimental design in OECD aquatic toxicology test guidelines.* Shell Research Ltd., Sittingbourne Research Centre, Sittingbourne, Kent, R.-U. 42. p.

———. 1998. A discussion of the NOEC/ANOVA approach to data analysis. Document de travail, 9 p., dans : *OECD, 1998. Report of the OECD workshop on statistical analysis of aquatic toxicity data. OECD, Paris.* OECD Environmental Health and Safety Publications, Series on Testing and Assessment nº 10, 133 p.

Paine, M.D., 1996. Repeated measures designs. Lettre au directeur de la publication. *Environ. Toxicol. Chem,. 13:*1439–1441.

———. 2002. Statistical significance in environmental effects monitoring (EEM) programs. *SETAC Globe, 3 (1):* 23–24. [Society of Environmental Toxicology and Chemistry, Pennsacola, Floride.]

Parmar, M.K.B. et D. Machin, 1995. *Survival analysis:Aa practical approach.* Wiley and Sons, New York.

Pickering, W., J. Lazorchak et K. Winks, 1996. Subchronic sensitivity of one-, four-, and seven-day old fathead minnow *(Pimephales promelas)* larvae to five toxicants. *Environ. Toxicol. Chem., 15:*353–359.

Porebski, L.M. et J.M.Osborne, 1998. The application of a tiered testing approach to the management of dredged sediments for disposal at sea in Canada. *Chemistry and Ecology, 14:*197–214.

Rand, G.M. (dir.), 1995. *Fundamentals of aquatic toxicology: effects, environmental fate, and risk assessment.* 2e éd. Taylor & Francis, Washington, D.C., 1125 p.

Rand, G.M. et S.R. Petrocelli (dir.), 1985. *Fundamentals of aquatic toxicology.* Hemisphere Pub., Washington, D.C.

Ricker, W.E., 1958. Handbook of computations for biological statistics of fish populations. *Bull. Fish. Res. Bd Canada,* no 119, 300 p.

Robertson, J.L., K.C. Smith, N.E. Savin et J.L. Lavigne, 1984. Effects of dose selection and sample size on the precision of lethal dose estimates in dose-mortality regression. *J. Econ. Entomol., 77:*883–837.

Rosner, B., 1983. Percentage points for a generalized ESD many-outlier procedure. *Technometrics, 25:*165–172.

Rowe, D.W., J.B. Sprague, T.A. Heming et I.T. Brown, 1983. Sublethal effects of treated liquid effluent from a petroleum refinery. II. Growth of rainbow trout. *Aquat. Toxicology, 3:*161–169.

Salsburg, D., 2001. *The lady tasting tea. How statistics revolutionized science in the twentieth century.* Henry Holt & Co., New York, 340 p.

SAS [SAS Institute Inc.], 1988. *SAS procedures guide,* version 6.03, et *Additional SAS/STAT procedures,* version 6.03 (rapport technique P-179 de SAS). SAS Institute Inc., Cary, Caroline du Nord. [http://www.sas.com]

———. 2000. SAS/STAT users guide, version 9, SAS Institute Inc., Cary.

Scholze, M., W. Boedeker, M. Faust, T. Backhaus, R. Altenburger et L.H. Grimme, 2001. A general best-fit method for concentration-response curves and the estimation of low-effect concentrations. *Environ. Toxicol. Chem., 20:*448–457.

Schroeder, J. et R.P. Scroggins, 2001. Meeting notes. Discussion of comments on the fourth draft version of guidance document on statistical methods to determine endpoints of toxicity tests. 27 et 28 sept. 2001, Centre des sciences environnementales du Pacifique, North Vancouver, C.-B.

Scott, D.W., 1992. *Multivariate density estimation. Theory, practice and visualization.* Wiley and Sons, New York.

Scroggins, R.P., J.A. Miller, A.I. Borgmann et J.B. Sprague, 2002. Sublethal toxicity findings by the pulp and paper industry for Cycles 1 and 2 of the environmental effects monitoring program. *Water Qual. Res. J. Canada, 37:(1):*21–48.

Searle, S.R., 1971. *Linear models.* John Wiley & Sons, New York.

Sebaugh, J.L., 1998. Comparison of LC50 results from commonly used computer programs. p. 383-397, dans : *Environmental toxicology and risk assessment: seventh volume.* E.E. Little, A.J. DeLonay et B.M. Greenberg (dir.), ASTM STP 1333, Amer. Soc. Testing and Materials, Philadelphie, 416 p.

Shapiro, S.S. et M.B. Wilk, 1965. An analysis of variance test for normality (complete samples). *Biometrika, 52:*591–611.

Shepard, M.P., 1955. Resistance and tolerance of young speckled trout *(Salvelinus fontinalis)* to oxygen lack, with special reference to low oxygen acclimation. *J. Fish. Res. Board Can., 12:*387–446.

Shirley, E., 1977. A non-parametric equivalent of Williams' test for contrasting increasing dose levels of a treatment. *Biometrics, 33:*386–389.

Shukla, R., W. Wang, F. Fulk, C. Deng et D. Denton, 2000. Bioequivalence approach for whole effluent toxicity testing. *Environ. Toxicol. Chem., 19:*169–174.

Slob, W., 2002. Dose-response modelling of continuous endpoints. *Toxicol. Sc., 66:*298–312.

Snedecor, G.W. et W.G. Cochran, 1980. *Statistical methods.* 7e éd. Iowa State Univ. Press, Ames, Iowa.

Sokal, R.R. et F.J. Rohlf, 1981. *Biometry.* W.H. Freeman and Co., San Francisco, Calif.

Sprague, J.B., 1964. Lethal concentrations of copper and zinc for young Atlantic salmon. *J. Fish. Res. Board Can. 21:*17–26.

————. 1969. Measurement of pollutant toxicity to fish -- I. Bioassay methods for acute toxicity. *Water Res., 3:* 793–821.

————. 1995. Factors that modify toxicity. p. 1012–1051, dans : *Fundamentals of aquatic toxicology.* G.M. Rand (dir.). Taylor and Francis, Washington, D.C.

———————. 1997. Review of methods for sublethal aquatic toxicity tests relevant to the Canadian metal-mining industry. Natural Resources Canada, Aquatic Effects Technol. Eval. Progr., Ottawa, Ont. AETE Project 1.2.1: 102 p.

Sprague, J.B. et A. Fogels, 1977. Watch the Y in bioassay. Proc. 3rd. Aquatic Toxicity Workshop, Halifax, 2–3 nov. 1976. Environnement Canada, rapport de surveillance EPS-5-AR-77-1: 107-118.

SPSS, 1996. SPSS 6.1 for Windows. SPSS Inc., 233 South Wacker Drive, Chicago, Ill. 60606-5307 [conçu pour Windows 3.1. comprend les régressions probit et logit].

———. 2001. SPSS base 11.0 for Windows. SPSS Inc., 233 South Wacker Drive, Chicago, Ill. 60606-5307 [les régressions probit et logit font partie du module « Régression », que l'on peut ajouter au progiciel de statistique de base].

Steel, R.G.D., 1959. A multiple comparison rank sum test: treatments versus control. *Biometrics, 15:*560–572.

———————. 1960. A rank-sum test for comparing all pairs of treatments. *Technometrics, 2:*197–611.

———————. 1961. Some rank sum multiple comparison tests. *Biometrics, 17:*539–552.

Steel, R.G.D. et J.H. Torrie, 1980. *Principles and procedures of statistics.* 2e éd. McGraw-Hill Book Co., New York.

Steel, R.G.D., J.H. Torrie et D.A. Dickey, 1997. *Principles and procedures of statistics: a biometrical approach.* 3e éd. McGraw-Hill Book Co., Boston.
666 p.

Stephan, C.E., 1977. Methods for calculating an LC_{50}. p. 65–84, dans : *Aquatic toxicology and hazard evaluation.* F.L. Mayer et J.L. Hamelink (dir.), Amer. Soc. Testing and Materials, Philadelphie. ASTM STP no 634.

Stephan, C.E., K.A. Busch, R. Smith, J. Burke et R.W. Andrew, 1978. A computer program for calculating an LC50 [LC50.BAS]. Fourni à titre gracieux par C.E. Stephan, U.S. Environmental Protection Agency, Duluth, Minn.

Stephan, C.E. et J.W. Rogers, 1985. Advantages of using regression analysis to calculate results of chronic toxicity endpoints. p. 328–338, dans : *Aquatic toxicology and hazard assessment: Eighth symposium.* R.C. Bahner et D.J. Hansen (dir.). Amer. Soc. Testing and Materials, Philadelphie. ASTM STP 891.

Stephan, C.E., D.I. Mount, D.J. Hansen, J.H. Gentile, G.A. Chapman et W.A. Brungs, 1985. Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses. USEPA, Office of Research and Development, Environmental Research Laboratories, Washington, D.C. [accès sous le no PB 85-227049 du National Technical Information Service (NTIS)].

Stephenson, G.L., N. Koper, G.F. Atkinson, K.R. Solomon et R.P. Scroggins, 2000. Use of nonlinear regression techniques for describing concentration-response relationships of plant species exposed to contaminated site soils. *Environ. Toxicol. Chem., 19:*2968–2981.

Suter, G.W. II, 1996. Abuse of hypothesis testing statistics in ecological risk assessment. *Human and Ecol. Risk Assess., 2:*331–347.

Suter, G.W. II, A.E. Rosen, E. Linder et D.F. Parkhurst, 1987. Endpoints for responses of fish to chronic toxic exposures. *Environ. Toxicol. Chem., 6:*793–809.

Suter, G.W. II, B.W. Cornaby, C.T. Hadden, R.N. Hull, M. Stack et F.A. Zafran, 1995. An approach for balancing health and ecological risks at hazardous waste sites. *Risk. Anal., 15:*221–231.

SYSTAT, 1990. SYSTAT: the system for statistics. SYSTAT Inc., Evanston, Illinois. 677 p.

Thompson, W.R., 1947. Use of moving averages and interpolation to estimate median-effective dose. 1. Fundamental formulas, estimation of error, and relation to other methods. *Bact. Reviews, 11:*115–145.

Tattersfield, F. et H.M. Morris, 1924. An apparatus for testing the toxic values of contact insecticides under controlled conditions. *Bull. Entomological Res., 14:* 223–233.

TOXCALC. Version 5.0, 1994. Tidepool Scientific Software, McKinleyville, Calif. 95521. [Programme sur disquette et guide imprimé de l'utilisateur. Remplacés, en 2001, par le logiciel *CETIS*. (v. ce mot.)] [http://members.aol.com/tidesoft/toxcalc]

TOXSTAT, 1996. Version 3.5. Lincoln Research Associates, Inc., P.O. Box 4276, Bisbee, Ariz., 85603, courriel : danlra@msn.com. [Programmes sur disquette et guide imprimé de l'utilisateur]

Tukey, J.W., 1977. *Exploratory data analysis*. Addison-Wesley, Reading, Mass. 688 p.

USEPA [United States Environmental Protection Agency], 1991. Technical support document for water quality-based toxics control. USEPA, Office of Water, Washington, D.C., EPA/505/2-90-001.

———. 1994a. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to freshwater

organisms. 3e éd. U.S. EPA, Environmental Monitoring Systems Laboratory, Cincinnati, Ohio, EPA 600/4-91-002.

———. 1994b. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to marine and estuarine organisms. 2nd ed. USEPA, Environmental Monitoring Systems Laboratory, Cincinnati, Ohio, EPA 600/4-91/003.

———. 1994c. Methods for measuring the toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates. USEPA, Duluth, Minn., EPA/600/R-94/024.

———. 1994d. Methods for assessing the toxicity of sediment-associated contaminants with estuarine and marine amphipods. USEPA, Office of Research and Development, Washington, D.C. EPA/600/R-94/025.

———. 1995. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to west coast marine and estuarine organisms. G.A. Chapman, D.L. Denton et J.M. Lazorchak (dir.), USEPA, Office of Research and Development, Washington, D.C., EPA 600/R-95/136, 661 p.

———. 2000a. Method guidance and recommendations for whole effluent toxicity (WET) testing (40 CFR Part 136). USEPA, Office of Water, Washington, D.C., EPA 821-B-00-004. 60 p.

———. 2000b. Understanding and accounting for method variability in whole effluent toxicity applications under the national pollutant discharge elimination system program. USEPA, Office of Wastewater Management Washington, D.C., EPA/833/R-00/003.

USEPA et USACE [United States Environmental Protection Agency et United States Army Corps of Engineers], 1994. Evaluation of dredged material proposed for discharge in inland and near coastal waters. USEPA, Office of Science and Technology, Washington, D.C., EPA/000/0-93/000.

van der Hoeven, N., 1991. LC_{50} estimates and their confidence intervals derived for tests with only one concentration with partial effect. *Water Res., 25:*401–408

———. 1997. How to measure no effect. Part III: Statistical aspects of NOEC, ECx and NEC estimates. *Environmetrics, 8:*255–261.

van der Hoeven, N., F. Noppert et A. Leopold, 1997. How to measure no effect. Part I: Towards a new measure of chronic toxicity in ecotoxicology. Introduction and workshop results. *Environmetrics, 8:*241–248.

Van Ewijk, P.H. et J.A. Hoekstra, 1993. Calculation of the EC50 and its confidence interval when subtoxic stimulus is present. *Ecotox. Env. Safety 25:*25–32.

Villeneuve, D.L., A.L. Blankenship et J.P. Giesy, 2000. Derivation and application of relative potency estimates based on in vitro bioassay results. *Environ. Toxicol. Chem., 19:*2835–2843.

Wang, Q., D.L. Denton et R. Shukla, 2000. Applications and statistical properties of minimum significant difference-based criterion testing in a toxicity testing program. *Environ. Toxicol. Chem., 19:*113–117.

Wang, S.C.D. et E.P. Smith, 2000. Adjusting for mortallity effects in chronic toxicity testing: mixture model approach. *Environ. Toxicol. Chem., 19:*204–209.

Wardlaw, A.C., 1985. *Practical statistics for experimental biologists.* John Wiley & Sons, Toronto, Ont..

Warren, C.E. 1971. *Biology and control of water pollution.* Saunders, Toronto, Ont. 434 p.

Wellek, S., 2002. *Testing statistical hypotheses of equivalence.* Chapman & Hall/CRC, Boca Raton, Floride. 290 p.

WEST, Inc. et D.D. Gulley, 1996. Toxstat® 3.5. Western EcoSystems Technology, Inc., Cheyenne, Wyoming, U.S.A. [logiciel et mode d'emploi]

Wilber, C.G., 1962. The biology of water toxicants in sublethal concentrations. p. 326–331, dans : *Biological problems in water pollution. Third seminar.* C.M. Tarzwell (dir.), U.S. Public Health Service, Dept. Health, Education, and Welfare, R.A. Taft Sanitary Engineering Center, Cincinnati, Ohio, P.H.S. pub. no 999-WP-25.

Williams, D.A., 1971, A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics, 27:*103–117.

———. 1972. The comparison of several dose levels with a zero dose control. *Biometrics, 28:*519–531.

WSDOE [Washington State Dept of Ecology], 1998. Laboratory guidance and whole effluent toxicity test review criteria. WSDOE, Water Quality Program, pub. no WQ-R-95-80, 71 p. Olympia, Washington.

Zajdlik, B.A., 1996. An introduction to threshold modelling of non-quantal bioassay data. p. 89–96, dans : *Proc. 22nd Ann. Aquat. Toxicity Workshop: October 2-4, 1995, St. Andrews, New Brunswick.* K. Haya et A.J. Niimi (dir.), Fisheries and Oceans, Can. Tech. Rept Fisheries and Aquatic Sc. no 2093.

———. (en préparation) Methods for statistically comparing EC50s and ICps. B. Zajdlik & Associates Inc., Rockwood, Ont.

Zajdlik, B.A., K.G. Doe et L.M. Porebski, 2000. Report on biological toxicity tests using pollution gradient studies -- Sydney Harbour. Environnement Canada, Service de la protection de l'environnement, Division du milieu marin, EPS 3/AT/2. 104 p.

Zajdlik, B.A., G. Gilron, P. Riebel et G. Atkinson, 2001. The $500,000 fish. *SETAC Globe, 2 (1):* 28–30. [Society of Environmental Toxicology and Chemistry, Pennsacola, Floride.]

Zaleski, R.T., G.E. Bragin, M.J. Nicolich, W.R. Arnold et A.L. Middleton, 1997. Comparison of growth endpoint estimation methods in EPA effluent short-term chronic testing guidelines. Affiche PWA088, Soc. Environ. Toxicology and Chemistry, 18th Annual Meeting, San Francisco, 16-20 nov. 1997.

Zar, J.H., 1974. *Biostatistical analysis.* Prentice-Hall, Inc., Englewood Cliffs, N.J.

———. 1999. *Biostatistical analysis*. 4e éd. Prentice-Hall, Inc., Upper Saddle River, N.J.

Annexe A

Méthodes d'essai biologique et guides à l'appui, publiés par la Section de l'élaboration et de l'application des méthodes d'Environnement Canada [1]

Titre de la méthode ou du guide	Type de données	Date de publication [2]	Date de modification
A. — Méthodes génériques (universelles)			
Essai de létalité aiguë sur la truite arc-en-ciel [SPE 1/RM/9]	Quantiques : mortalité aiguë	juillet 1990 (1990a)	mai 1996
Essai de létalité aiguë sur l'épinoche à trois épines (*Gasterosteus aculeatus*) [SPE 1/RM/10]	Quantiques : mortalité aiguë	juillet 1990 (1990b)	mars 2000
Essai de létalité aiguë sur *Daphnia spp.* [SPE 1/RM/11]	Quantiques : mortalité aiguë	juillet 1990 (1990c)	mai 1996
Essai de reproduction et de survie sur le cladocère *Ceriodaphnia dubia* [SPE 1/RM/21]	Effet double : mortalité des adultes et nombre de jeunes	févr. 1992 (1992a)	nov. 1997
Essai de croissance et de survie sur des larves de tête-de-boule [SPE 1/RM/22]	Effet double : mortalité et poids des larves	févr. 1992 (1992b)	nov. 1997
Essai de toxicité sur la bactérie luminescente *Photobacterium phosphoreum* [maintenant *Vibrio fischeri*] [SPE 1/RM/24]	Quantitatives : inhibition à 50 % de la luminescence	nov. 1992 (1992c)	—
Essai d'inhibition de la croissance de l'algue d'eau douce *Selenastrum capricornutum* [maintenant *Pseudokirchneriella subcapitata*] (SPE 1/RM/25)	Quantitatives : pourcentage spécifié de réduction de la production de cellules algales pendant 72 h	nov. 1992 (1992d)	nov. 1997
Essai de toxicité aiguë de sédiments chez les amphipodes marins ou estuariens [SPE 1/RM/26]	Quantiques : pourcentage de survie, émergence du sédiment, pas de fouissement ultérieur	déc. 1992 (1992e)	oct. 1998
Méthode d'essai biologique : essai sur la fécondation chez les échinides (oursins verts et oursins plats) [SPE 1/RM/27]	Quantiques : diminution du taux de fécondation	déc. 1992 (1992f)	nov. 1997
Méthode d'essai biologique : essais toxicologiques sur des salmonidés (truite arc-en-ciel) aux premiers stades de leur cycle biologique [SPE 1/RM/28, 2e édition]	Quantiques : embryons, alevins ou truitelles non viables ; mortalité des truitelles. Quantitatives : poids des truitelles. Description de toute manifestation d'un développement retardé ou anormal.	juillet 1998 (1998a)	—
Méthode d'essai biologique : essai de survie et de croissance des larves dulcicoles de chironomes (*Chironomus tentans* ou *Chironomus riparius*) dans les sédiments [SPE 1/RM/32]	Effet double : survie et poids des larves	déc. 1997 (1997a)	—
Méthode d'essai biologique : essai de survie et de croissance de l'amphipode dulcicole *Hyalella azteca* dans les sédiments [SPE 1/RM/33]	Effet double : survie et poids des larves	déc. 1997 (1997b)	—

1. On peut acheter ces documents des Publications de la Protection de l'environnement, Environnement Canada, Ottawa, K1A 0H3. Pour obtenir de plus amples renseignements ou formuler des observations, prière de s'adresser au chef de la Division des méthodes biologiques, Centre de technologie environnementale, Environnement Canada, Ottawa, K1A 0H3.
2. À la date de publication s'ajoute le code utilisé sous la rubrique « Références » (par ex. 1990a).

Titre de la méthode ou du guide	Type de données	Date de publication [2]	Date de modification
Méthode d'essai biologique : essai de mesure de l'inhibition de la croissance de la plante macroscopique dulcicole *Lemna minor* [SPE 1/RM/37]	Effet double : poids et diminution de la prolifération (du nombre) des frondes	mars 1999 (1999b)	—
Méthode d'essai biologique. Essai de survie et de croissance des vers polychètes spionides (*Polydora cornuta*) dans les sédiments [SPE 1/RM/41]	Effet double : survie et poids	déc. 2001 (2001a)	—
Essais pour déterminer la toxicité de sols contaminés pour les vers de terre *Eisenia andrei, Eisenia fetida* ou *Lumbricus terrestris* [SPE 1/RM/43]	Quantiques : mortalité aiguë, taux d'évitement. Effet double : mortalité d'adultes, nombre et poids de jeunes	juin 2004 (2004a)	—
Essais de mesure de la levée et de la croissance de plantes terrestres exposées à des contaminants dans le sol [SPE 1/RM/45]	Effet double : nombre de semis ayant levé, longueur et poids des pousses et des racines	juin 2004 (2004b)	—
Essai de mesure de la survie et de la reproduction de collemboles exposés à des contaminants dans le sol [SPE 1/RM/47]	Effet double : survie des adultes et nombre de jeunes	déc. 2007 (2007)	—
B. — Méthodes de référence [3]			
Méthode de référence pour la détermination de la létalité aiguë d'effluents chez la truite arc-en-ciel [SPE 1/RM/13, 2e édition]	Quantiques : mortalité aiguë	déc. 2000 (2000a)	—
Méthode de référence pour la détermination de la létalité aiguë d'effluents chez *Daphnia magna* [SPE 1/RM/14, 2e édition]	Quantiques : mortalité aiguë	déc. 2000 (2000b)	—
Méthode d'essai biologique : méthode de référence pour la détermination de la létalité aiguë d'un sédiment pour des amphipodes marins ou estuariens [SPE 1/RM/35]	Quantiques : survie à court terme	déc. 1998 (1998b)	—
Méthode de référence servant à déterminer la toxicité des sédiments à l'aide d'une bactérie luminescente dans en un essai en phase solide [SPE 1/RM/42]	Quantitatives : inhibition de la luminescence	avril 2002 (2002a)	—
C. — Guides			
Document d'orientation sur le contrôle de la précision des essais de toxicité au moyen de produits toxiques de référence [SPE 1/RM/12]		août 1990 (1990d)	—
Document d'orientation sur le prélèvement et la préparation de sédiments en vue de leur caractérisation physico-chimique et d'essais biologiques [SPE 1/RM/29]		déc. 1994 (1994)	—
Document d'orientation sur la mesure de la précision des essais de toxicité sur sédiments dopés avec un produit toxique de référence [SPE 1/RM/30]		sept. 1995	—
Guide des essais écotoxicologiques employant une seule espèce et de l'interprétation de leurs résultats [SPE 1/RM/34]		déc. 1999 (1999a)	—
Guide des essais de pathogénicité et de toxicité de nouvelles substances microbiennes pour les organismes aquatiques et terrestres [SPE 1/RM/44]		mars 2004 (2004d)	—
Document d'orientation sur les méthodes statistiques applicables aux essais d'écotoxicité [SPE 1/RM/46] (le présent document)		mars 2005	—

3. On entend par *méthode de référence* une méthode biologique particulière d'essai de toxicité, assortie d'un ensemble de consignes et de conditions décrites avec précision dans un document écrit. Contrairement aux méthodes génériques d'essai biologique d'Environnement Canada, les méthodes de référence sont habituellement associées à des dispositions réglementaires précises.

Annexe B

Composition du Groupe intergouvernemental sur l'écotoxicité (en janvier 2004)

Administration fédérale, Environnement Canada

C. Blaise
Centre Saint-Laurent
Montréal

M. Bombardier
Centre de technologie environnementale
Ottawa

U. Borgmann
Institut national de recherche sur les eaux
Burlington (Ontario)

J. Bruno
Centre des sciences environnementales du Pacifique (CSEP)
North Vancouver (Colombie-Britannique)

C. Buday
CSEP
North Vancouver

K. Doe
Centre des sciences de l'environnement (CSE) de la région de l'Atlantique
Moncton

G. Elliott
Service de la protection de l'environnement
Edmonton

F. Gagné
Centre Saint-Laurent
Montréal

M. Harwood
Service de la protection de l'environnement
Montréal

D. Hughes
CSE, région de l'Atlantique
Moncton

P. Jackman
CSE, région de l'Atlantique
Moncton

N. Kruper
Service de la protection de l'environnement
Edmonton

M. Linssen
CSEP
North Vancouver

D. MacGregor
Centre de technologie environnementale
Ottawa

L. Porebski
Direction du milieu marin
Gatineau (Québec)

J. Princz
Centre de technologie environnementale
Ottawa

G. Schroeder
CSEP
North Vancouver

R.P. Scroggins
Centre de technologie environnementale
Ottawa

T. Steeves
CSE, région de l'Atlantique
Moncton

D. Taillefer
Direction du milieu marin
Gatineau

S. Trottier
Centre Saint-Laurent
Montréal

G. van Aggelen (président)
CSEP
North Vancouver

B. Walker
Centre Saint-Laurent
Montréal

P.G. Wells
Service de la conservation de l'environnement
Dartmouth (Nouvelle-Écosse)

Administration fédérale, Pêches et Océans Canada

R. Roy
Institut Maurice-Lamontagne
Mont-Joli (Québec)

Administration fédérale, Ressources naturelles Canada

J. McGeer
Laboratoire des sciences minérales, Centre canadien de la technologie des minéraux et de l'énergie (CANMET)
Ottawa

B. Vigneault
Laboratoire des sciences minérales, CANMET
Ottawa

J. Beyak
Laboratoire des sciences minérales, CANMET
Ottawa

Administrations provinciales

C. Bastien
ministère de l'Environnement du Québec
Sainte-Foy

B. Bayer
Environnement Manitoba
Winnipeg

M. Mueller
ministère de l'Environnement de l'Ontario
Rexdale

D. Poirier
ministère de l'Environnement de l'Ontario
Rexdale

J. Schroeder
ministère de l'Environnement de l'Ontario
Rexdale

T. Watson-Leung
ministère de l'Environnement de l'Ontario
Rexdale

Administration centrale et bureaux régionaux d'Environnement Canada

Administration centrale
351, boul. Saint-Joseph
Place Vincent-Massey
Gatineau (Québec)
K1A 0H3

Région de l'Ontario
4905, rue Dufferin, 2e étage
Downsview
M3H 5T4

Région de l'Atlantique
15e étage, Queen Square
45, Alderney Drive
Dartmouth, Nouvelle-Écosse
B2Y 2N6

Région des Prairies et du Nord
pièce 210, Twin Atria No. 2
4999, 98e avenue
Edmonton, Alberta
T6B 2X3

Région du Québec
105, rue McGill
8e étage
Montréal
H2Y 2E7

Région du Pacifique et du Yukon
401, rue Burrard
Vancouver
V6C 3S5

Calculs employant des concentrations arithmétiques et logarithmiques

D.1 Exemple : comparaison de moyennes

Dans le tableau ci-dessous, nous exposons les écarts entre les médianes, les moyennes arithmétiques et les moyennes géométriques ou logarithmiques de quatre ensembles hypothétiques de nombres. Les colonnes pourraient renfermer des chiffres pouvant représenter les CE 50 estimées à la faveur d'essais répétés. La première colonne représente de « bonnes » données, les résultats étant assez semblables les uns aux autres. L'ensemble de données de la colonne B comporte une valeur élevée, légèrement divergente. Celui de la 3e colonne possède une concentration des plus improbables. L'ensemble de la 4e colonne comporte une concentration aberrante extrêmement peu probable. Pour les besoins de l'exemple, posons qu'il n'y a aucune raison de rejeter une concentration. Tout principe général dégagé à la faveur des exemples extrêmes s'appliquerait aux ensembles ordinaires de données des laboratoires de toxicologie.

	« Bonnes données »	Valeur divergente	Valeur peu probable	Données bizarres
	10	10	10	10
	12	12	12	12
	14	14	14	14
	16	16	16	16
	18	18	18	18
	22	28	100	1 000
Médiane	15	15	15	15
Moyenne arithmétique	15,3	16,3	28	178
Moyenne géométrique	14,6	15,4	19	28

Pour le « bon » ensemble, celui de la 1re colonne, les trois mesures de la tendance centrale sont essentiellement les mêmes, comme on s'y attendrait de données régulières. Dans les trois colonnes suivantes, la médiane reste la même, parce qu'elle ne prend pas en considération la valeur numérique de l'élément dont la valeur est maximale. La médiane pourrait souvent être un bon choix pour exprimer la tendance centrale d'une distribution asymétrique. En effet, quand on estime la CE 50, la base de cette estimation est l'effet quantique exercé sur l'organisme *médian.* Cependant, dans d'autres secteurs de la toxicologie, on a rarement trouvé à employer la médiane dans des notions quantitatives telles que la concentration, car les chercheurs privilégient plutôt la moyenne, qui utilise les valeurs numériques. Dans les exemples qui précèdent, la médiane ne parvient pas à révéler l'aberrance d'une valeur élevée ; même si les deux valeurs maximales de l'ensemble étaient anormalement élevées, la médiane ne changerait pas.

La moyenne arithmétique de la 2e colonne est supérieure de 6 % à la moyenne géométrique. C'est une différence notable, mais sans grande importance.

Dans le 3e ensemble, qualifié de « peu probable », la moyenne arithmétique est près de 1,5 fois supérieure à la géométrique, ce qui est appréciable. La moyenne géométrique tend à diminuer l'effet de la valeur aberrante et elle est plus représentative des cinq autres concentrations qui se suivent et sont rapprochées les unes des autres.

Dans le dernier exemple, celui des données qualifiées de bizarres, la moyenne arithmétique est 5,4 fois plus grande que la moyenne géométrique et elle n'est absolument pas représentative des valeurs de la série. La moyenne géométrique est, du moins, du même ordre de grandeur que les cinq concentrations semblables.

Normalement, les valeurs aberrantes des deux colonnes de droite pourraient être rejetées après application d'un test statistique, mais ce n'est pas le but de l'exemple. Dans les deux cas extrêmes, la moyenne géométrique offre clairement une défense plus robuste que la moyenne arithmétique contre les concentrations exceptionnellement fortes et elle semblerait mieux représenter la toxicité moyenne probable. Le principe étant établi, il s'appliquerait également aux ensembles de « bonnes » données. La moyenne géométrique devrait représenter de façon plus fiable les valeurs moyennes. Les lecteurs pourraient imaginer d'autres exemples.

D.2 Exemple : les régressions probit

Le tableau qui suit donne les résultats des calculs de CE 50 par régression probit. Les quatre exemples correspondent aux ensembles de données énumérés dans le tableau 2 du § 4.4. Les estimations des CE 50 employant les logarithmes de la concentration sont celles que l'on obtient grâce à la plupart des programmes informatiques, qui utilisent automatiquement le logarithme de la concentration dans le calcul. Les estimations des CE 50 employant les concentrations arithmétiques ont été obtenues à l'aide du programme TOXSTAT 3.5, en évitant l'emploi de logarithmes. (Cette erreur serait facile à faire dans ce programme, sans jamais s'en apercevoir, ce qui est une bonne raison pour vérifier les estimations au moyen de graphiques tracés à la main.)

	CE 50 (et leurs limites de confiance) pour quatre ensembles exemplatifs de données			
	A	B	C	D
Avec la conc. arithmétique	6,3 (4,9–7,7)	20,6 (14,3–26,9)	15,6 (11,4–19,5)	32,5 (17,6–47,4)
Avec la conc. logarithmique	5,6 (4,4–7,2)	16,8 (12,1–23,3)	12,8 (9,4–17,6)	26,5 (13,3–53,1)
Quotient de la conc. arithmétique par la logarithmique	1,12	1,23	1,22	1,23

Les CE 50 calculées avec les concentrations arithmétiques sont en moyenne 1,2 fois plus élevées que les valeurs appropriées. C'est une erreur appréciable, que l'on devrait éviter. La plupart des intervalles de confiance sont également déplacées vers le haut. Le § 4.4 et le tableau 2 montrent que, lorsque l'on utilise les concentrations logarithmiques appropriées, les paramètres de toxicité calculés à l'aide de TOXSTAT concordent, sur l'essentiel, avec les calculs effectués par d'autres programmes.

La randomisation

La randomisation intervient dans la répartition des organismes d'expérience entre les récipients et les concentrations et dans la répartition des récipients dans le dispositif expérimental.

E.1 Nombres aléatoires pour la répartition des organismes entre les récipients

La randomisation de la répartition des organismes entre les récipients n'est pas exigée par toutes les méthodes publiées par Environnement Canada. On a jugé que, dans certains essais, cette méthode risquait de causer des erreurs plus graves de manipulation. On a plutôt jugé que la randomisation subséquente des récipients ou des concentrations suffirait pour éviter la commission d'erreurs systématiques (biais) dans l'essai et dans ses résultats.

Cependant, si on peut manipuler les organismes d'expérience comme des individus (par ex. les poissons, comme dans l'exemple présenté ci-dessous) et si on peut les compter dans les enceintes expérimentales, il est toujours avantageux de le faire aléatoirement. On pourrait utiliser toute méthode commode telle que le tirage, d'un chapeau, de bouts de papier marqués des concentrations utilisées. La plupart des ordinateurs peuvent produire des nombres aléatoires. Une autre façon commode est offerte par l'USEPA (1995), au moyen d'une table de nombres aléatoires. Nous la reproduisons ici (table E.1).

On commence par affecter aux diverses concentrations expérimentales des nombres à deux chiffres, que l'on dispose dans un tableau comme celui qui suit immédiatement ci-dessous. On pourrait utiliser plusieurs nombres à deux chiffres pour chaque concentration, de sorte que, ultérieurement, on aura utilisé la totalité des nombres dans une table de nombres aléatoires. On n'utilise pas la valeur 00 et, dans le tableau ci-dessous, aucun nombre n'est supérieur à 30.

Nombres attribués					Concentration expérimentale
01	07	13	19	25	Témoin
02	08	14	20	26	0,5 % d'effluent
03	09	15	21	27	1 % d'effluent
04	10	16	22	28	2,5 % d'effluent
05	11	17	23	29	5 % d'effluent
06	12	18	24	30	10 % d'effluent

Maintenant, à partir d'une table de nombres aléatoires telle que la table E.1, on peut y choisir n'importe quelle rangée et colonne pour débuter (par ex. la rangée 3 de la colonne 6, qui correspond à la valeur 19). On attribue ce nombre au premier poisson que l'on retire du vivier et, d'après le tableau ci-dessus, le poisson est attribué au groupe témoin.

De retour à la table E.1, on choisit le deuxième poisson en se déplaçant horizontalement vers la droite ; on ignore les nombres 64, 50 et 93, supérieurs aux nombres utilisés dans le tableau qui précède. Le deuxième poisson correspond au nombre 03, qui est attribué à la concentration 1 % d'effluent. On poursuit le choix des nombres dans la rangée de la table E.1, puis dans la rangée suivante, jusqu'à ce que l'on ait rempli les réservoirs, disons avec 10 poissons chacun. Il faut consigner les affectations des poissons pour que chaque réservoir reçoive son plein effectif de poissons, mais sans sujet excédentaire. Si on tire un nombre qui ferait en sorte que le réservoir renfermerait trop de poissons, on ignore ce nombre. Si une seule personne effectue cette opération, il est plus facile de choisir les nombres sur papier, puis de capturer et de répartir les poissons.

Table E.1. — Nombres aléatoires à deux chiffres (USEPA, 1995, d'après Dixon et Massey, 1983)

10 09 73 25 33	76 52 01 35 86	34 67 35 43 76	80 95 90 91 17	39 29 27 49 45
37 54 20 48 05	64 89 47 42 96	24 80 52 40 37	20 63 61 04 02	00 82 29 16 65
08 42 26 89 53	19 64 50 93 03	23 20 90 25 60	15 95 33 47 64	35 08 03 36 06
99 01 90 25 29	09 37 67 07 15	38 31 13 11 65	88 67 67 43 97	04 43 62 76 59
12 80 79 99 70	80 15 73 61 47	64 03 23 66 53	98 95 11 68 77	12 27 17 68 33
66 06 57 47 17	34 07 27 68 50	36 69 73 61 70	65 81 33 98 85	11 19 92 91 70
31 06 01 08 05	45 57 18 24 06	35 30 34 26 14	86 79 90 74 39	23 40 30 97 32
85 26 97 76 02	02 05 16 56 92	68 66 57 48 18	73 05 38 52 47	18 62 38 85 79
63 57 33 21 35	05 32 54 70 48	90 55 35 75 48	28 46 82 87 09	83 49 12 56 24
73 79 64 57 53	03 52 96 47 78	35 80 83 42 82	60 93 52 03 44	35 27 38 84 35
98 52 01 77 67	14 90 56 86 07	22 10 94 05 58	60 97 09 34 33	50 50 07 39 98
11 80 50 54 31	39 80 82 77 32	50 72 56 82 48	29 40 52 42 01	52 77 56 78 51
83 45 29 96 34	06 28 89 80 83	13 74 67 00 78	18 47 54 06 10	68 71 17 78 17
88 68 54 02 00	86 50 75 84 01	36 76 66 79 51	90 36 47 64 93	29 60 91 10 62
99 59 46 73 48	87 51 76 49 69	91 82 60 89 28	93 78 56 13 68	23 47 83 41 13
65 48 11 76 74	17 46 85 09 50	58 04 77 69 74	73 03 95 71 86	40 21 81 65 44
80 12 43 56 35	17 72 70 80 15	45 31 82 23 74	21 11 57 82 53	14 38 55 37 63
74 35 09 98 17	77 40 27 72 14	43 23 60 02 10	45 52 16 42 37	96 28 60 26 55
69 91 62 68 03	66 25 22 91 48	36 93 68 72 03	76 62 11 39 90	94 40 05 64 18
09 89 32 05 05	14 22 56 85 14	46 42 75 67 88	96 29 77 88 22	54 38 21 45 98
91 49 91 45 23	68 47 92 76 86	46 16 28 35 54	94 75 08 99 23	37 08 92 00 48
80 33 69 45 98	26 94 03 68 58	70 29 73 41 35	53 14 03 33 40	42 05 08 23 41
44 10 48 19 49	85 15 74 79 54	32 97 92 65 75	57 60 04 08 81	22 22 20 64 13
12 55 07 37 42	11 10 00 20 40	12 86 07 46 97	96 64 48 94 39	28 70 72 58 15
63 60 64 93 29	16 50 53 44 84	40 21 95 25 63	43 65 17 70 82	07 20 73 17 90
61 19 69 04 46	26 45 74 77 74	51 92 43 37 29	65 39 45 95 93	42 58 26 05 27
15 47 44 52 66	95 27 07 99 53	59 36 78 38 48	82 39 61 01 18	33 21 15 94 66
94 55 72 85 73	67 89 75 43 87	54 62 24 44 31	91 19 04 25 92	92 92 74 59 73
42 48 11 62 13	97 34 40 87 21	16 86 84 87 67	03 07 11 20 59	25 70 14 66 70
23 52 37 83 17	73 20 88 98 37	68 93 59 14 16	26 25 22 96 63	05 52 28 25 62
04 49 35 24 94	75 24 63 38 24	45 86 25 10 25	61 96 27 93 35	65 33 71 24 72
00 54 99 76 54	64 05 18 81 59	96 11 96 38 96	54 69 28 23 91	23 28 72 95 29
35 96 31 53 07	26 89 80 93 45	33 35 13 54 62	77 97 45 00 24	90 10 33 93 33
59 80 80 83 91	45 42 72 68 42	83 60 94 97 00	13 02 12 48 92	78 56 52 01 06
46 05 88 52 36	01 39 09 22 86	77 28 14 40 77	93 91 08 36 47	70 61 74 29 41
32 17 90 05 97	87 37 92 52 41	05 56 70 70 07	86 74 31 71 57	85 39 41 18 38
69 23 46 14 06	20 11 74 52 04	15 95 66 00 00	18 74 39 24 23	97 11 89 63 38
19 56 54 14 30	01 75 87 53 79	40 41 92 15 85	66 67 43 68 06	84 96 28 52 07
45 15 51 49 38	19 47 60 72 46	43 66 79 45 43	59 04 79 00 33	20 82 66 95 41
94 86 43 19 94	36 16 81 08 51	34 88 88 15 53	01 54 03 54 56	05 01 45 11 76
98 08 62 48 26	45 24 02 84 04	44 99 90 88 96	39 09 47 34 07	35 44 13 18 80
33 18 51 62 32	41 94 15 09 49	89 43 54 85 81	88 69 54 19 94	37 54 87 30 43
80 95 10 04 06	96 38 27 07 74	20 15 12 33 87	25 01 62 52 98	94 62 46 11 71
79 75 24 91 40	71 96 12 82 96	69 86 10 25 91	74 85 22 05 39	00 38 75 95 79
18 63 33 25 37	98 14 50 65 71	31 01 02 46 74	05 45 56 14 27	77 93 89 19 36
74 02 94 39 02	77 55 73 22 70	97 79 01 71 19	52 52 75 80 21	80 81 45 17 48
54 17 84 56 11	80 99 33 71 43	05 33 51 29 69	56 12 71 92 55	36 04 09 03 24
11 66 44 98 83	52 07 98 48 27	59 38 17 15 39	09 97 33 34 40	88 46 12 33 56
48 32 47 79 28	31 24 96 47 10	02 29 53 68 70	32 30 75 75 46	15 02 00 99 94
69 07 49 41 38	87 63 79 19 76	35 58 40 44 01	10 51 82 16 15	01 84 87 69 38

Il faut recommencer à zéro la randomisation pour chaque essai ou ensemble d'essais. Il n'est pas approprié de toujours utiliser la même suite de nombres aléatoires.

E.2 Nombres aléatoires pour répartir la position des enceintes

Dans un essai, la position des récipients est importante. Dans une disposition rectiligne, l'une des extrémités pourrait se trouver près d'une fenêtre et recevoir directement un éclairage puissant, ce qui entraînerait une réaction de stress, des tentatives d'évitement ou une meilleure croissance, selon le type d'organisme. Une des extrémités pourrait être exposée à des températures supérieures à celles de l'autre extrémité, ce qui pose des risques particuliers si l'essai a été effectué dans un incubateur. Elle pourrait se trouver près d'une porte, et la circulation pourrait faire sursauter les organismes. Ces facteurs et d'autres que l'on ne soupçonne pas pourraient influer sur les résultats. L'expérimentateur devrait s'efforcer de supprimer ou de réduire au minimum ces influences, mais des facteurs méconnus pourraient jouer. Une façon de supprimer toute variable systématique non contrôlée consiste à randomiser les positions des enceintes expérimentales.

Dans l'exemple qui précède, il y avait cinq concentrations et un témoin. On pourrait les placer en une seule rangée, pour des raisons de commodité. Leurs positions pourraient être randomisées en tirant les nombres d'un chapeau ou, comme nous venons de le faire, en employant la table E.1 ou une table plus simple se trouvant dans un manuel de mathématique.

Si on employait plus de récipients, disons pour cinq répétitions de chaque concentration, on pourrait vouloir les disposer en 6 rangées de 5 sur la paillasse (quelle que soit la disposition, la même méthode s'appliquerait). On pourrait employer le même procédé de randomisation que celui que nous venons de décrire.

- On crée un tableau de 5 rangées sur 6 colonnes de nombres successifs de 01 à 30, représentant les 30 positions de la configuration.

Voici comme se présenterait le tableau des positions des enceintes expérimentales.

01	02	03	04	05	06
07	08	09	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30

- Dans la table E.1, débuter en tout point et lire le nombre qui s'y trouve ; il représentera la première répétition du témoin (par ex. la 11^{e} colonne de la table, 3^{e} rangée. On y lit le nombre 23, de sorte que la première répétition du témoin va à la position 23 du tableau qui précède, près de l'angle inférieur droit).

- Dans la table E.1, on va vers le prochain chiffre, à droite de celui que l'on vient de choisir, et ce nombre correspond à la 2^{e} répétition du témoin (c'est le nombre 20, de sorte que cette répétition sera placée près de l'angle inférieur gauche du tableau).

- Poursuivre jusqu'à ce que l'on ait attribué une place à toutes les répétitions de toutes les concentrations. Si un nombre est tiré une seconde fois, on n'en tient pas compte, tout comme si le nombre excède 30. (Dans notre exemple, le nombre ensuite tiré serait 90, qui ne s'applique à aucune position, de sorte qu'on n'en tient pas compte et que le nombre ensuite tiré, 25, devient la position de la 3^{e} répétition du témoin.)

Dans le tableau qui précède et qui renferme des nombres de 01 à 30, on pourrait avoir inscrit deux ou trois nombres pour chaque position, si on avait voulu, afin de mieux épuiser les nombres de la table E.1 et de ne pas avoir à tirer tant de nombres qu'il fallait ignorer. L'emploi d'un seul nombre pour chaque position signifie simplement que l'expérimentateur tirera surtout de la table des nombres inutilisables pour les positions et dont il ne faudra pas tenir compte.

La disposition établie de cette façon serait une préparation parfaite pour un ensemble d'observations en aveugle. On consignerait les positions des diverses répétitions, mais en ne les faisant pas connaître de l'observateur pendant l'essai. De la sorte, l'observateur n'introduirait pas de biais du fait de sa connaissance des concentrations. Après la fin de l'essai, on apparierait les observations à leurs répétions et concentrations appropriées.

Exception possible. — L'essai de substances volatiles pourrait être l'une des rares situations dans laquelle la randomisation des enceintes expérimentales dans un dispositif expérimental ouvert ne serait pas approprié. Le toxique volatil pourrait s'échapper des récipients qui en renferment de fortes concentrations et parvenir aux autres récipients. Cela pourrait notamment contaminer les témoins et provoquer des effets anormaux chez ces derniers. Bien que cela puisse certainement influer sur l'analyse statistique, le remède se trouve dans un autre domaine, celui des installations de laboratoire convenablement conçues pour les essais de matières volatiles. Une telle situation exigerait des récipients scellés, une ventilation séparée ou quelque autre solution du genre. La randomisation des traitements resterait un objectif.

Calcul de la moyenne et des limites sur une carte de contrôle

Pour des raisons de commodité, nous répétons la fig. 2 dans la présente annexe, où elle devient la fig. F.1. Les données de cette figure serviront à montrer les étapes des calculs dans les cartes de contrôle. Nous présentons les étapes avec un certain détail, parce que les expérimentateurs d'aujourd'hui risquent de ne pas être à l'aise dans l'emploi des logarithmes. Une feuille de calcul facilite les calculs. Elle se chargera des logarithmes, des antilogarithmes et elle calculera la moyenne et l'écart type. D'ailleurs, les calculs sont assez simples à faire sur une calculette scientifique.

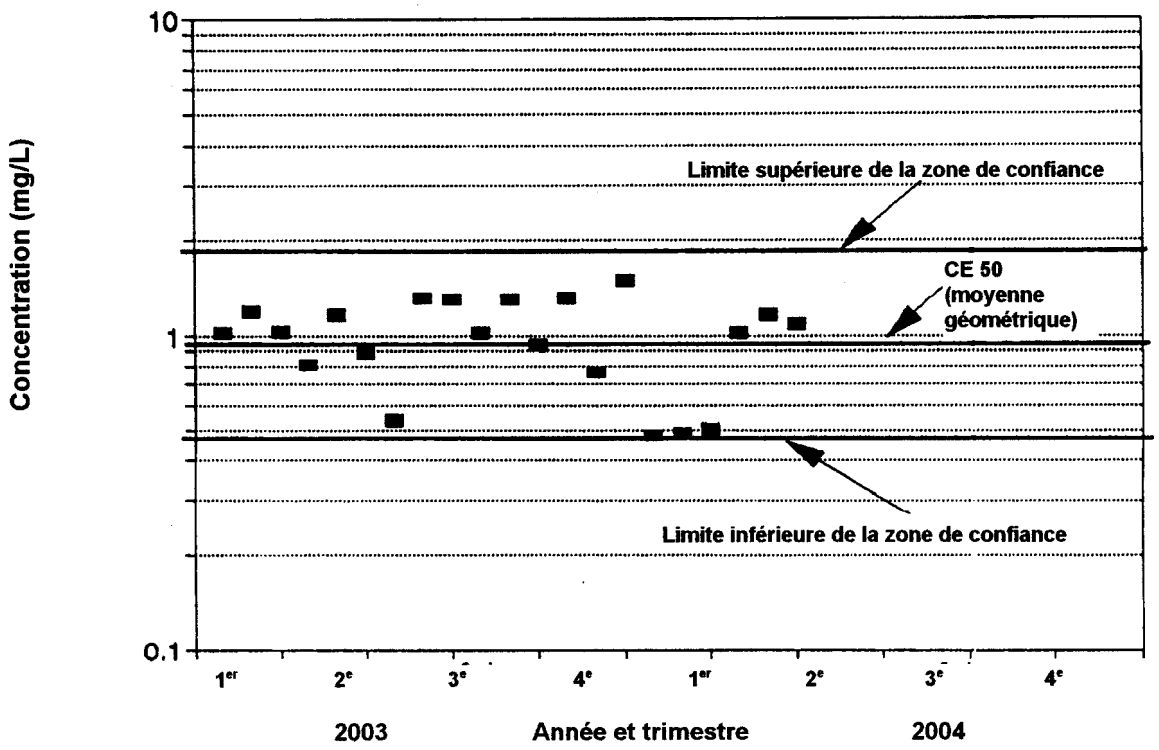


Figure F.1. — Carte de contrôle pour les essais avec un toxique de référence. Cette carte montre les résultats d'essais d'un toxique de référence en milieu aquatique, effectués dans un laboratoire canadien.

Les étapes du calcul de la zone de confiance sont comme suit :

(1) Compiler les données antérieures. Il s'agit des CE 50 estimées pour le toxique de référence au laboratoire. Comme ces dernières seraient probablement consignées sous leur valeur arithmétique, il faudrait les convertir en logarithmes. (Le logarithme de base 10 est usuel, bien que les logarithmes népériens soient également bons si on les utilise constamment.) Pour les besoins de l'exemple, nous n'énumérons que les cinq premières CE 50 de la fig. F.1 et, pour les logarithmes, nous ne précisons qu'une partie des chiffres.

Antilogarithme de la CE 50	1,02	1,19	1,03	0,81	1,16
Logarithme de la CE 50	0,0086002...	0,075547...	0,012837...	– 0,091515...	0,064458...

(2) On fait la moyenne des logarithmes. C'est simplement la *moyenne* arithmétique des logarithmes. La moyenne des 21 logarithmes des CE 50 de la fig. F.1 est – 0,027356... Cette moyenne reste sous forme logarithmique pour les calculs ultérieurs, mais elle est plus intelligible sous sa forme arithmétique, qui est de 0,93895 mg/L (avant de l'arrondir). La valeur arithmétique de 0,94 est la moyenne géométrique des CE 50. Dans la fig. F.1, on la représente par une droite.

(3) On calcule l'écart type des 21 logarithmes des CE 50. Il vaut 0,15288...

(4) La valeur de deux écarts types est deux fois la valeur calculée à l'étape 3 : 2 × 0,15288... = 0,30576... L'antilogarithme vaut (avant arrondissement) 2,0219, mais il n'est d'aucune utilité particulière.

(5) La limite supérieure de la zone de confiance se calcule comme la moyenne (étape 2) plus 2 écarts types (étape 4) : – 0,02736... + 0,30576... = 0,278404...
On peut convertir cette valeur en son antilogarithme de 1,9 mg/L, limite supérieure de la zone de confiance et que l'on peut représenter sur la carte de contrôle (v. la fig. F.1).

C'est une erreur de faire les calculs des étapes 5 et 6 avec des valeurs arithmétiques ; ils donnent des réponses erronées. (Voir cependant le texte qui suit pour utiliser les valeurs arithmétiques avec la multiplication et la division plutôt que l'addition et la soustraction.)

(6) On calcule la limite inférieure de la zone de confiance comme la moyenne (étape 2) moins 2 écarts types (étape 4) : – 0,02736... – 0,30576... = – 0,33312...
Convertie en son antilogarithme de 0,46 mg/L, cette limite inférieure de la zone de confiance est portée sur le graphique (fig. F.1.)

Dans la fig. F.1, les limites de la zone de confiance sont symétriques par rapport à la moyenne parce que l'axe vertical est une échelle logarithmique. Par le passé, des investigateurs peu à l'aise avec les notions de logarithme ont été contrariés par le fait que les limites de la zone de confiance calculées de la façon qui précède n'étaient pas symétrique quand on les portait sur une échelle arithmétique. On ne devrait pas s'inquiéter de cela. **Les limites bien calculées ne seront jamais symétriques sur une échelle arithmétique (elles ne le devraient pas) elles seront symétriques sur une échelle logarithmique appropriée.**

Il y a une autre façon de calculer les limites de confiance, si on le souhaite, à l'aide de valeurs arithmétiques. L'addition et la soustraction de logarithmes correspondent respectivement à la multiplication et à la division de leurs équivalents arithmétiques.

- Ainsi, la limite supérieure de confiance pourrait être calculée comme la moyenne géométrique (étape 2) multipliée par l'antilogarithme de deux écarts types (étape 4) :
  0,938954 × 2,0219 = 1,9 mg/L, la même valeur que celle nous avons obtenue à l'étape 5.

- On pourrait estimer la limite inférieure de confiance comme étant la moyenne géométrique divisée par l'antilogarithme de deux écarts types :
  0,938954 / 2,0219 = 0,46 mg/L, de nouveau la valeur obtenue à l'étape 6.

Il existe aussi une autre façon de représenter graphiquement les données. On pourrait employer une échelle arithmétique pour l'axe vertical et on pourrait représenter les valeurs logarithmiques. La plupart des expérimentateurs considéreraient

probablement cela comme plus encombrant. Les tableurs ont simplifié la représentation des valeurs sur un graphique avec une échelle logarithmique.

Il vaut la peine de comparer les limites erronées de la zone de confiance que l'on aurait obtenues si on n'avait pas employé de logarithmes dans les calculs, c'est-à-dire si les calculs s'étaient fondés sur les valeurs arithmétiques des CE 50 : 1,02, 1,19, 1,03, 0,81, etc.

- La moyenne ainsi calculée aurait été de 0,99 mg/L, ce qui est quelque peu plus élevé que la valeur appropriée de 0,94 mg/L.

- Les limites de confiance auraient été de 1,6, au lieu de 1,9 mg/L, et de 0,39, au lieu de 0,46 mg/L. Ainsi, sur la fig. F.1., elles auraient été sensiblement abaissées. L'écart entre les limites aurait été plus petit, à 1,2 mg/L au lieu de 1,4 mg/L.

À prime abord, il pourrait sembler anormal que la moyenne erronée soit plus élevée que la moyenne logarithmique, alors que les limites erronées de la zone de confiance sont plus basses que celle que l'on calcule logarithmiquement. Cela est une caractéristique prévisible de la distorsion. Les limites (arithmétiques) erronées sont également espacées au-dessus et au-dessous de la moyenne sur une échelle arithmétique. Les limites de confiance calculées logarithmiquement ne sont pas équidistantes de leur moyenne sur une échelle arithmétique, mais elles sont, comme il se doit, symétriques en tant que multiples de la moyenne arithmétique, n'en différant que d'un facteur d'environ 2.0.

Variation « convenable » des CE 50. — Comme nous l'avons mentionné dans le § 2.8.1, Environnement Canada s'est dit d'avis que la variation, dans des essais répétés d'un toxique de référence, serait considérée comme acceptable si le coefficient de variation était 30 % et même, de préférence, de 20 %. Ce nombre-guide résulte de calculs effectués avec des paramètres arithmétiques de toxicité, méthode susceptible d'être biaisée et à proscrire. On a donc modifié logarithmiquement le nombre-guide, d'une manière approximative, par le procédé exposé ci-dessous. Dans les calculs, on a employé des chiffres significatifs supplémentaires, et, dans le texte qui suit, les points de suspension suivant une valeur logarithmique signifient que l'on en a omis les chiffres qui en feraient normalement partie.

On a compilé plusieurs ensembles réels et factices de CE 50. À l'aide des valeurs arithmétiques des CE 50, on a calculé les coefficients de variation. On a ajusté l'un des ensembles de CE 50 de façon à ce que le coefficient de variation égalât 30,0 % et un autre ensemble de façon à ce qu'il égalât 20,0 %. Ensuite, pour chaque ensemble de CE 50, on a calculé l'écart type à l'aide des valeurs logarithmiques des CE 50. Les coefficients de variation arithmétiques et les écarts types logarithmiques ont révélé une relation approximativement rectiligne quand on les a portés sur un graphique. On a sélectionné, dans la relation, les écarts types logarithmiques pour qu'ils correspondent aux coefficients de variation arithmétiques de 30 et de 20 %.

Les écarts types étaient de 0,132... et de 0,0338… et ils représentent la traduction de la règle empirique d'Environnement Canada concernant une variation « raisonnable » (acceptable) et « préférable » dans un ensemble de résultats. Les mêmes valeurs s'appliquent à tout ensemble de résultats, parce qu'elles ont été obtenues à partir de rapports sur une échelle logarithmique. Les écarts types réels (calculés) pour tout ensemble de CE 50 logarithmiques peuvent être comparés à ces nombres-guides.

Pour les données de la fig. F.1, on peut calculer un écart type réel de 0,153, ce qui est supérieur à la valeur acceptable de 0,132 estimée antérieurement. On peut conclure que les données de la figure sont quelque peu plus variables que le nombre-guide acceptable qu'Environnement Canada a publié.

(Si un ensemble de données possédant la même moyenne que les données de la fig. F.1 avait effectivement un écart type égal au nombre-guide de 0,132, la zone de confiance serait quelque peu plus étroite que la zone analogue de la figure.

Les limites seraient de ± 2 écarts types par rapport à la moyenne. La moyenne logarithmique étant de − 0,027356... (voir ci-dessus), les limites seraient de − 0,0273 56 − (2 × 0,132) et de − 0,027356 + (2 × 0,132), c'est-à-dire de − 0,2914... et de 0,2366..., dont les antilogarithmes sont respectivement de 0,51 et de 1,7 mg/L. Ces limites, pour des données hypothétiques, sont un peu plus rapprochées de la moyenne que les limites réelles de la zone de confiance montrées dans la fig. F.1, qui sont de 0,46 et 1,9 mg/L.)

[Si, pour un ensemble de données possédant la même moyenne que les données de la fig. F.1, l'écart type égalait le nombre-guide de 0,0338, la zone de confiance serait encore plus étroite. Par des calculs analogues à ceux de l'alinéa qui précède, on arriverait à des limites, pour la zone de confiance, de 0,80 et de 1,1 mg/L.)

Ces règles empiriques pour déterminer une variation acceptable et préférable dans les résultats d'essais répétés de toxicité pourraient être perçues comme étant quelque peu optimistes.

Tests s'appliquant aux résultats d'essai à concentration unique, sans répétition

Les essais dont il est question sont habituellement assujettis à un programme réglementaire tel que la surveillance des rejets de déchets. Un critère rigoureux du type « réussite ou échec » permet de juger des résultats de l'essai ; cependant, on pourrait avoir besoin d'appliquer des tests statistiques. Il sera question ici, pour compléter l'information donnée dans la section 3, de certains de ces tests.

G.1 Méthode exacte de Fisher

On peut appliquer la méthode exacte de Fisher à un seul échantillon et témoin, sans répétition. Souvent, l'effet observé est la mortalité, de sorte que les données sont quantiques. La méthode exacte de Fisher, appliquée uniquement aux données quantiques permet de comparer les résultats. C'est un test unilatéral de signification statistique, parce que l'expérimentateur veut savoir si la mortalité est plus grande dans l'échantillon, que chez le témoin. Ce test pourrait aussi convenir à l'égard de la mortalité constatée dans les essais ayant employé *Ceriodaphnia* (EC, 1992a).

On peut voir la méthode en action dans l'exemple qui suit, montrant le nombre d'organismes dans un essai à un seul échantillon, sans répétition.

	Morts	Vivants	Total
Échantillon	6	4	10
Témoin	1	9	10
Total	7	13	20

Le nombre d'organismes vivants et morts est présenté dans le tableau qui précède. L'hypothèse nulle est que la proportion de morts n'est pas plus grande dans l'échantillon que chez le témoin.

Dans les marges droite et inférieure du tableau, on totalise chaque rangée et colonne. On appelle n le total général du tableau. En l'occurrence, $n = 20$.

On choisit le plus petit des quatre totaux partiels, en l'occurrence 7, que l'on désigne par m_1. Dans l'autre marge du tableau (où ne se trouve pas m_1), on choisit le plus petit total, que l'on appelle m_2. Ici, $m_2 = 10$; on choisit le 10 inférieur (correspondant au témoin) ; on arriverait au même résultat quel que soit le nombre 10 choisi. La sélection suivante porte sur les nombres du corps du tableau qui contribuent à la fois à m_1 et à m_2. C'est 1, que l'on peut appeler f.

L'étape suivante consiste à comparer f avec des valeurs critiques se trouvant dans une table plutôt complexe, exposée dans certains manuels de statistique tels que celui de Zar (1999 ; « *Critical values for Fisher's exact test* » [valeurs critiques pour la méthode exacte de Fisher] »). On consulte la table à un certain point, conformément au niveau choisi de signification (habituellement une probabilité de 0,05) et, également, d'après les valeurs de n, m_1 et m_2. À cet emplacement de la table, on trouvera deux paires de valeurs critiques de f, et on devrait utiliser la première paire, qui est destinée à un test unilatéral. (La seconde paire est destinée à un test bilatéral, qui n'entre pas dans notre propos). Si f est inférieur ou égal à la première valeur critique ou est supérieur ou égal à la seconde valeur critique, on rejette l'hypothèse nulle et on conclut que la mortalité est plus grande dans l'échantillon que chez le témoin.

Dans le cas qui nous occupe, les valeurs critiques tirées de la table sont 1 et 6. Comme la valeur calculée de f est 1, f est égal à la première valeur critique ; on rejette donc l'hypothèse nulle, et on conclut que l'échantillon a présenté une

mortalité significativement accrue. (Si on avait choisi pour être m_2 le 10 de la rangée supérieure du tableau, *f* aurait été égal à 6, la deuxième valeur critique, ce qui aurait conduit aussi au rejet de l'hypothèse nulle.)

La mortalité dans le groupe expérimental, qui est de 6 sur 10 d'après le tableau, se trouve à être la plus faible mortalité qui serait significative pour un petit effet observé chez le témoin de 1 sur 10. Si, dans l'échantillon, le nombre de morts avait été de 5 sur 10, on aurait accepté l'hypothèse nulle. Cela n'est pas entièrement en désaccord avec la conclusion de Zajdlik *et al.* (2001) selon qui une décision du type réussite ou échec est ambiguë quand 4 à 7 poissons sur 10 meurent. S'il n'y avait pas d'effet observé chez les témoins, des mortalités moindres que 6 sur 10, à savoir 5 sur 10 et 4 sur 10, seraient significatives. Si l'effet observé chez le témoin était plus grand, disons de 2 morts sur 10, il faudrait des mortalités plus fortes (au moins 7 sur 10) pour qu'elles soient significatives.

Dans le § G.2, nous exposons une méthode de rechange, employant des diagrammes et des tableaux et fondée sur les tables de Finney.

G.2 Comparaison avec les tables de Finney

On peut comparer la mortalité dans un groupe à celle du témoin, à l'aide des diagrammes de la fig. G.1 ou des tables de Finney *et al.* (1963) dont ils dérivent. Les diagrammes montrés sont conçus pour 3 à 10 individus par groupe. Ils sont fournis par Wardlaw (1985), dans un manuel de statistique très à la portée des non-statisticiens, et ils ne fonctionnent que si les organismes sont en nombre égaux dans le groupe expérimental et le groupe témoin.

La fig. G.1 peut servir à tester l'exemple précédent, dans lequel la mortalité était de 6 sur 10 dans le groupe expérimental et 1 sur 10 chez le témoin. Les diagrammes sont conçus pour un test unilatéral de signification, de sorte que l'hypothèse nulle est que la mortalité dans le groupe expérimental n'est pas plus grande que chez le témoin.

Pour la comparaison de deux fois 10 individus, on emploie le diagramme inférieur droit (10 × 10). Pour la consultation du diagramme, on ne se sert que des numérateurs, c'est-à-dire de 6 pour le groupe expérimental et de 1 pour le groupe témoin. On se positionne donc dans la colonne 6 (mortalité dans le groupe expérimental) et dans la rangée 1 (mortalité chez le témoin). La case à l'intersection de la colonne et de la rangée est pointillée, ce qui signifie que la probabilité de cette survenue par le seul effet du hasard est de 0,05 ou moins. À ce niveau de probabilité, on rejette l'hypothèse nulle et on conclut que la mortalité dans le groupe expérimental est supérieure à mortalité chez le témoin. (À noter que la conclusion n'est pas que les deux groupes sont différents, ce qui implique une conclusion bilatérale selon laquelle la mortalité dans le groupe expérimental pourrait être soit plus grande, soit plus petite que chez le témoin.)

Les cases noires du diagramme correspondent aux combinaisons dont la probabilité est de 0,01 ou moins. Les cases blanches correspondent aux probabilités supérieures à 0,05, c'est-à-dire que la valeur du groupe expérimental n'excéderait pas significativement celle du groupe témoin par la valeur critique habituelle de *P*.

Pour les autres combinaisons que celles que montre la fig. G.1, on pourrait consulter les tables de Finney *et al.* (1963). Ces tables ne portent pas seulement sur les comparaisons de nombres égaux d'organismes, mais sur toutes les combinaisons possibles de nombres inégaux jusqu'à 40 par groupe. Par exemple, les tables permettent de comparer une mortalité de 18 sur 32 dans le groupe expérimental à une mortalité de 2 sur 20 dans le groupe témoin. Wardlaw (1985) explique aussi une méthode arithmétique fastidieuse de comparaison, qui deviendrait rapidement exorbitante dès que les effectifs des groupes dépasseraient 10 !

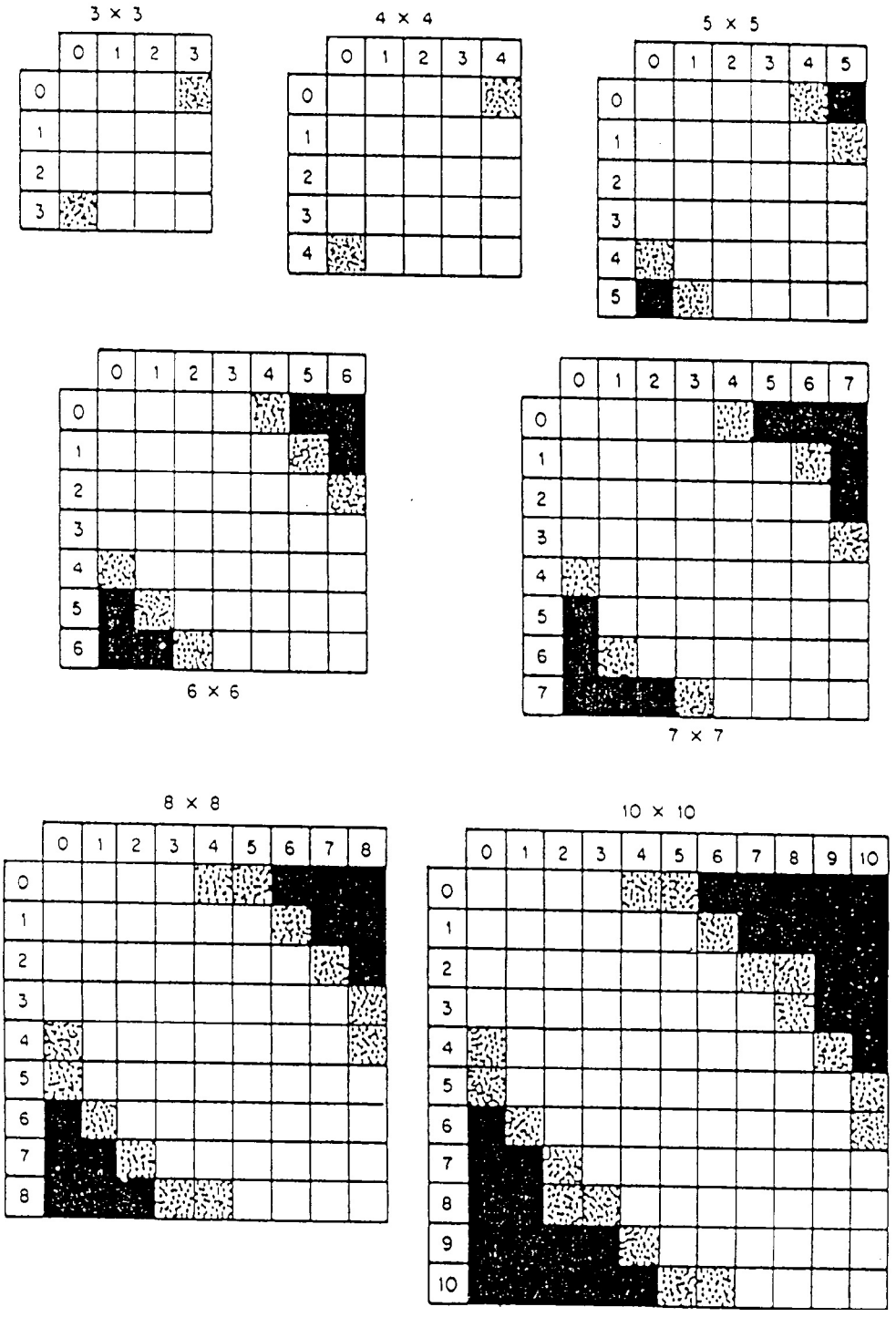


Figure G.1. — Diagrammes permettant la comparaison des effets quantiques dans un groupe expérimental et un groupe témoin. Les diagrammes permettent de déterminer si le groupe expérimental présente ou non un effet significativement plus grand que chez le groupe témoin. Ces diagrammes concernent un nombre égal d'unités expérimentales (organismes) dans les groupes expérimentaux et témoins, qui vont de 3 (diagramme supérieur gauche) à 10 (diagramme inférieur droit) chacun. Les probabilités dénotées par les cases noires, les cases pointillées et les cases blanches sont respectivement de $P \leq 1\ \%$; $5\ \% \geq P > 1\ \%$; $P > 5\ \%$. D'après Wardlaw (1985), à partir de diagrammes tirés de Finney *et al.* (1963).

G.3 Comparaison de deux proportions à l'aide d'un test Z

La méthode est expliquée dans les manuels généraux de statistique, habituellement sous l'appellation de « différences entre des proportions » ou de « comparaisons de proportions » (par ex. Zar, 1999 ; Snedecor et Cochran, 1980). Nous pouvons l'illustrer avec les données que nous avons utilisées pour la méthode exacte de Fisher.

	Morts	Vivants	Total	Proportion de morts
Groupe expérimental	6	4	10 = n_T	0,6 = p_T
Témoin	1	9	10 = n_C	0,1 = p_C
Total	7	13	20	
Proportion	0,35 = p_{TC}	0,65 = q_{TC}		

On peut calculer la statistique Z en introduisant les valeurs du tableau dans la formule suivante.

$$Z = \frac{p_T - p_C}{\sqrt{\frac{p_{TC}\, q_{TC}}{n_T} + \frac{p_{TC}\, q_{TC}}{n_C}}} = \frac{0,6 - 0,1}{\sqrt{\frac{0,35 \times 0,65}{10} + \frac{0,35 \times 0,65}{10}}} = \frac{0,5}{\sqrt{0,0455}} = 2,34 \qquad (G.1)$$

La valeur critique de Z pour $p = 0,05$ et un test unilatéral est la même que la valeur critique de t pour un nombre infini de degrés de liberté : 1,645. La valeur calculée de Z est supérieure à la valeur critique ; on rejette donc l'hypothèse nulle, et la mortalité dans l'enceinte expérimentale est supérieure à celle du témoin.

Explication de la notion de *probit* et de la transformation log-probit

H.1 Transformations usuelles

Les programmes informatiques de régression probit utilisent la transformation log-probit, que l'on voit dans les fig. 5, 8 et 9 du corps du texte. Cette transformation vise à redresser ce qui, sinon, serait une courbe normale cumulative (fig. H.1) asymétrique (v. le glossaire).

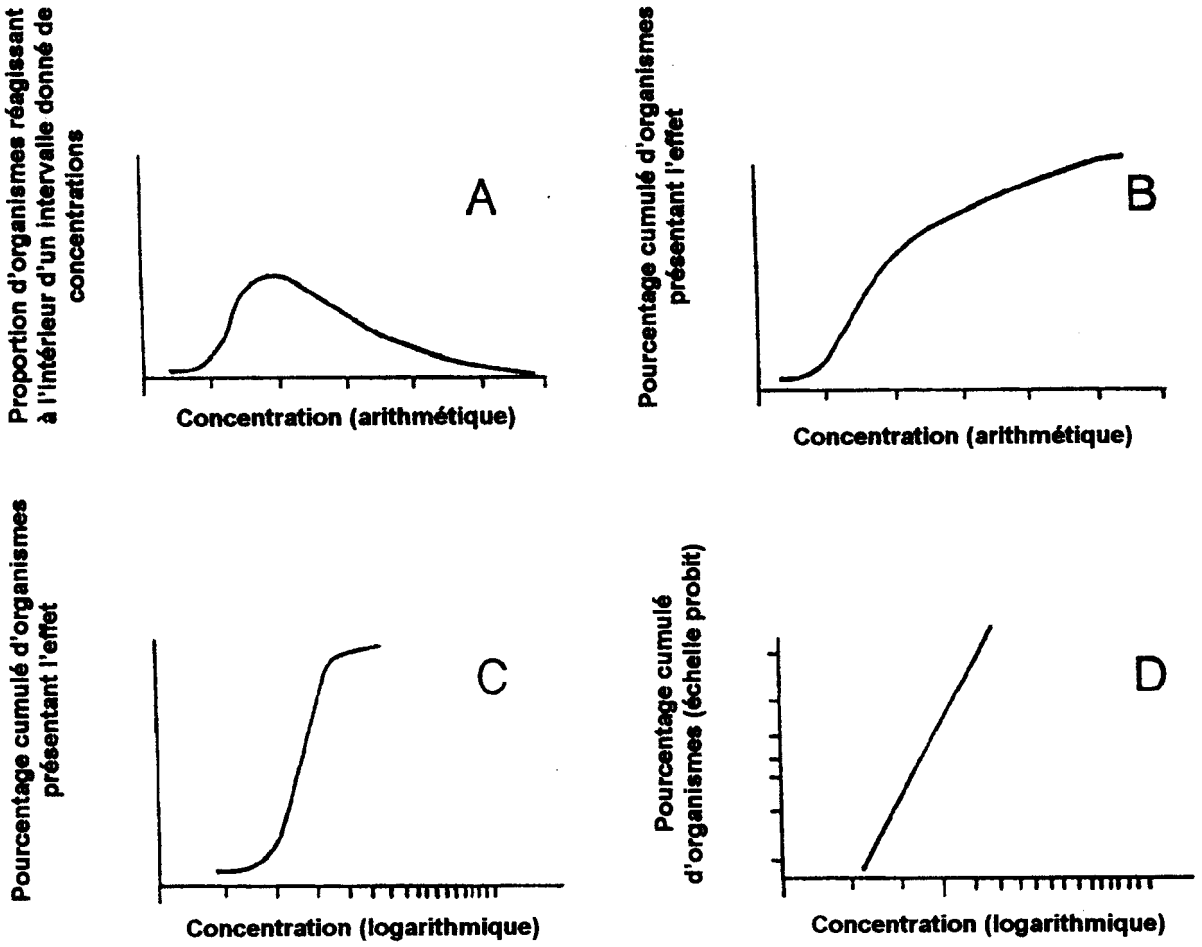


Figure H.1. — Transformation de données quantiques. Les résultats bruts d'un essai tel qu'un essai de toxicité létale pour le poisson donne habituellement une courbe normale asymétrique lorsqu'on porte les données sur un graphique dont les axes sont gradués arithmétiquement (graphique A). Cette distribution peut être cumulée pour donner une courbe sigmoïde asymétrique (graphique B). Le logarithme des concentrations supprime l'asymétrie (graphique C). L'application d'une transformation de probabilités au pourcentage d'effet (graphique D) redresse la courbe en comprimant verticalement sa partie centrale et en étirant progressivement ses extrémités, qui n'atteignent jamais, dans cette transformation, les valeurs de 0 ou de 100 %.

Si les résultats d'un essai de toxicité quantique étaient portés sur du papier réglé à échelle arithmétique, le résultat serait presque toujours une courbe normale asymétrique. Le graphique A de la fig. H.1 représente une telle courbe, de la proportion du nombre total d'organismes en expérience chez qui l'effet se serait manifesté à chaque concentration d'une suite d'intervalles de concentrations. Dans la partie gauche de la courbe, peu d'individus sont sensibles et manifestent l'effet à de faibles concentrations. Dans la partie droite, un nombre semblablement petit d'organismes est très résistant, ne manifestant l'effet qu'à de très fortes concentrations. La plupart des organismes sont touchés dans les intervalles médians de concentrations. Si on additionne le nombre d'organismes touchés, cela donne une courbe sigmoïde ou en S, asymétrique à droite (graphique B).

Un graphique des logarithmes de la concentration permet habituellement de supprimer l'asymétrie (graphique C). En utilisant une transformation de probabilité (= transformation en probits), on obtient une droite, montrée dans le graphique D. La droite permet l'application de techniques plus faciles d'ajustement de la distribution de données, ce qui a été important dans la mise au point de nouveaux modes opératoires et, par le passé, lorsque les calculs se faisaient à la main ou à l'aide d'une calculatrice mécanique. Aujourd'hui, grâce à l'ordinateur capable d'effectuer des calculs complexes, l'on pourrait se passer de la transformation en probits. Néanmoins, la vieille méthode usuelle de transformation log-probit continue d'être un bon modèle pour tracer un graphique à la main afin de vérifier la forme de la droite et la justesse des calculs informatiques.

H.2 Pourquoi des logarithmes ?

Dans un graphique fondé sur une échelle arithmétique des concentrations tel que la représentation des données brutes du graphique A de la fig. H.1, l'asymétrie à droite provient du fait qu'une augmentation arithmétique donnée représente successivement des proportions décroissantes de concentrations croissantes.

Une échelle logarithmique permet de résoudre convenablement ce problème de diminution des proportions, puisque l'augmentation, d'une proportion donnée d'une valeur arithmétique, n'importe laquelle (de 10 à 20, de 100 à 200 ou de 1 000 à 2 000) représente la même augmentation numérique d'un logarithme (§ 2.3). Ou, sur l'axe logarithmique d'un graphique, le doublement de la concentration occupe la même distance en valeur absolue, peu importe son point de départ sur l'axe. Cela est vrai des logarithmes de base 10 et, aussi, des logarithmes népériens de base *e.* On utilise systématiquement les logarithmes de base 10 en écotoxicologie, et il importe de ne pas mêler les types de logarithmes dans une analyse donnée.

H.3 Qu'est-ce qu'un probit ?

Les probits équivalent à l'écart type de la loi normale centrée réduite. À l'origine, en effet, on les a appelés *variables normales équivalentes* ou *NED* (pour *Normal Equivalent Deviate ;* Gaddum, 1953), expression signifiante pour les mathématiciens, mais aujourd'hui presque disparue. Dans l'analyse des données sur la toxicité quantique, les probits remplacent le pourcentage d'effet cumulé.

Les probits se fondent sur la distribution habituelle de fréquences d'une courbe normale : ± 1 écart type par rapport à la moyenne englobe environ 68 % des observations ; ± 2 écarts types englobent 95 % des observations ; etc. Si on dessine une courbe normale cumulative (sigmoïde), la relation théorique existant entre les pourcentages cumulés et les écarts types reste connue. Cette relation est utilisée avec les probits.

Un probit de valeur 1 (ou 1 probit) correspond à 1 écart type de la loi normale centrée réduite (distribution normale dont la moyenne égale 0 et la variance 1).

Plutôt qu'un exposé formel, on peut utiliser des diagrammes simplifiés pour montrer comment s'est dégagée la notion de probit (fig. H.2). Les graphiques sont expliqués dans les étapes qui suivent.

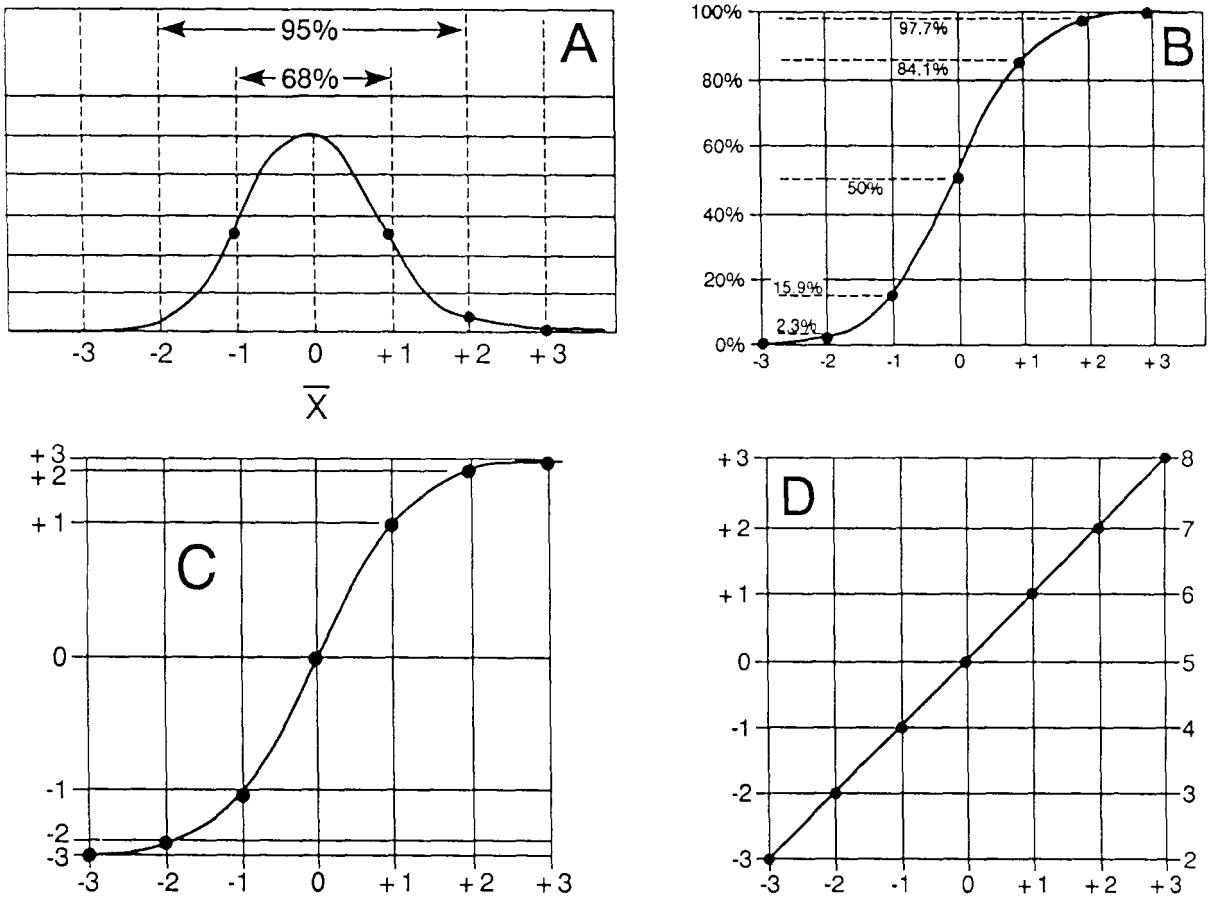


Figure H.2. — L'origine des probits. Voir l'explication des graphiques dans le texte.

(1) On commence par une courbe normale centrée réduite (graphique A de la fig. H.2). L'intervalle de ± 1 écart type par rapport à la moyenne englobe 68 % de la population (par définition de la courbe normale). L'intervalle de ± 2 σ englobe 95 % de la population et celui de ± 3 σ, 99,7 %, etc.

(2) On cumule la courbe. Les pourcentages se trouvent à fonctionner comme dans le graphique B de la fig. H.2, figurés sur les tiretés positionnés à diverses hauteurs sur le graphique. C'est une courbe sigmoïde typique.

(3) On supprime ensuite l'échelle des pourcentages sur l'axe vertical et on numérote les tiretés avec les mêmes nombres que ceux des ordonnées à l'origine sur l'axe horizontal (graphique C). Ces derniers nombres représentent les écarts types.

(4) L'échelle de la nouvelle numérotation de l'axe vertical du graphique C est irrégulière. On utilise à la place, une échelle arithmétique, qui va de – 3 à + 3 dans l'exemple du graphique D. On redresse ainsi la courbe sigmoïde. Si l'échelle des pourcentages avait subsisté, elle serait irrégulière, mais l'échelle fondée sur le nombre d'écarts types est régulière, et la courbe est devenue une droite.

Ce petit exercice dissipe le mystère. C'est simplement une méthode de redressement de la courbe normale cumulative. De pourcentages, les unités de l'axe vertical se sont transformées en équivalents d'écarts types, appelés à l'origine *variables normales équivalentes* et maintenant *probits*.

Une modification supplémentaire s'est imposée, et les expérimentateurs devraient en être conscients.

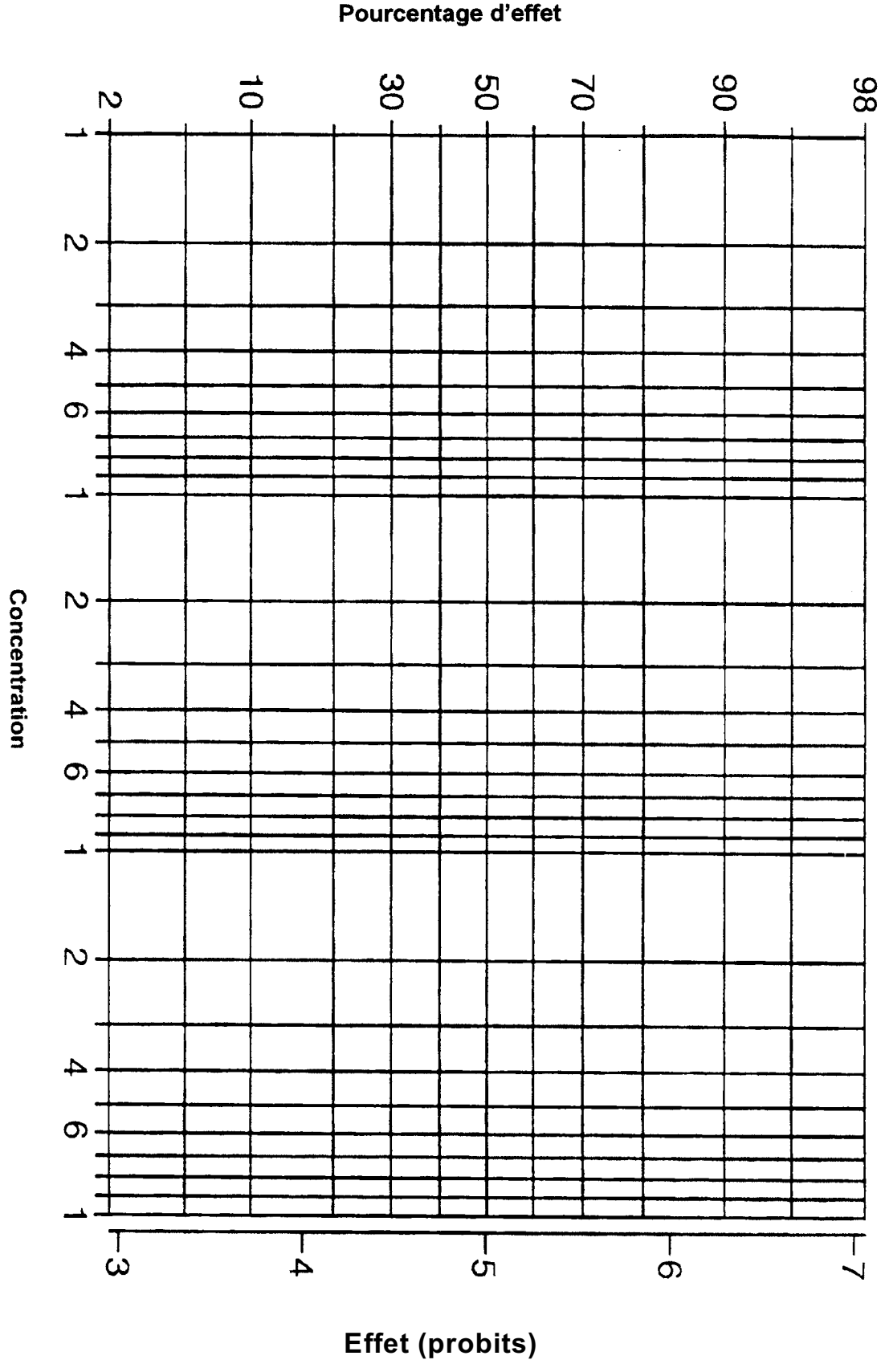
(5) L'échelle allant d'une valeur négative à une valeur positive était gênante à l'époque des calculettes. En conséquence, on a majoré chaque valeur de 5, de sorte que l'intervalle habituel de travail est devenu 2 à 8, comme il est montré sur la bordure de droite du graphique D. Ainsi, le probit 5 est devenu la médiane. À proprement parler, la définition de « probit » comprend la valeur ajoutée de 5. Pour les calculs à l'ordinateur, cette majoration n'est plus nécessaire, mais elle ne cause aucun tort.

Manifestement, il y a des relations entre les probits, les pourcentages et les écarts types d'une courbe normale. Les chercheurs peuvent donc sauter d'une notion à l'autre, s'ils le veulent. On peut trouver, dans des tables publiées, le probit de tout pourcentage particulier (Finney, 1971 ; Hubert, 1984 ; 1992), ou l'obtenir avec un calculateur de probabilité normale, que l'on trouve dans les feuilles calculs et les progiciels de statistique. Les programmes informatiques de régression probit font les calculs.

Papier log-probabilité (ou log-probit) vierge

À la page suivante, nous offrons un exemplaire de papier log-probit. On pourrait utiliser des photocopies de cette page pour les analyses, si ce type de papier est difficile à trouver. Ce papier permet de tracer le graphique des résultats d'essais de toxicité quantiques. L'effet est porté sur l'axe vertical. On pourrait y représenter tout effet quantique tel que la létalité, le taux de fécondation des œufs de salmonidés ou le pourcentage d'organismes présentant des lésions.

Sur le papier log-probit du commerce, on trouve divers axes. Sur certains, les échelles des probits vont de valeurs très petites à des valeurs très grandes (par ex. 0,1 % et 99,9 %), ce qui serait excessif la plupart du temps.


Pourcentage d'effet
2
10
30
50
70
90
98
Concentration
1
2
4
6
1
2
4
6
1
2
4
6
1
3
4
5
6
7
Effet (probits)

Avantages et explication des logits

Pour des motifs de simplicité mathématique et pour d'autres bonnes raisons, nous recommandons d'employer des méthodes logistiques plutôt que les probits. Cependant, les deux méthodes sont bonnes pour l'analyse des données quantiques, et, habituellement, les paramètres de toxicité qu'ils permettent d'estimer sont très semblables (§ 4.4).

L'analyse de données quantiques au moyen de logits est supérieure à l'analyse employant des probits, pour plusieurs raisons.

- Ces estimations sont numériquement plus stables que les estimations au moyen des probits ; elles sont moins susceptibles d'échouer (Hoekstra, 1989).

- Les paramètres obtenus par régression logistique utilisent toute l'information pertinente dans une série d'observations, ce qui n'est pas vrai de la régression probit. Inversement, les paramètres d'une régression logistique permettent de retrouver directement les données originelles.

- Les paramètres du modèle logistique sont largement utilisés comme mesures du risque dans les publications biomédicales.

- La programmation informatique des modèles de régression logistique est quelque peu plus facile.

- Des progiciels de statistique sont disponibles en nombre beaucoup plus grands que pour la régression probit.

Les tracés des fréquences cumulées des courbes normale et logit se ressemblent (fig. J.1). En conséquence, la transformation en logits peut donner des résultats satisfaisants avec les données qui suivent la loi normale et qui se prêtent à la régression probit. (Dans le § 4.5.1, nous décrivons comment l'effet binomial dans chaque enceinte expérimentale est devenu analysable à l'aide d'une distribution normale ou logistique, lorsque nous avons pris en considération la distribution cumulative pour toutes les enceintes.) La fig. J.1 montre que la courbe logistique a des ailes plus larges et plus « massives » que la courbe normale. Si l'expérimentateur est intéressé aux ailes de la courbe (disons la région inférieure à 5 % ou supérieure à 95 %), les paramètres de toxicité, estimés par les logits et les probits, seraient sensiblement différents.

Les courbes de la fig. J. 1 ont été centrées (autour d'une moyenne nulle) sur l'axe horizontal. Pour faciliter la compréhension, les données peuvent être traitées comme si elles étaient des poids d'organismes plutôt que comme des données quantiques. Ainsi la fig. J. 1 représente-t-elle les proportions cumulées d'organismes de poids divers. Pour centrer la courbe normale, on a soustrait le poids moyen de chaque observation individuelle de poids, puis on a divisé cette différence par l'écart type de l'ensemble de données. En conséquence, l'axe horizontal, sans unité, a simplement été légendé par *X*. Dans le cas de la courbe normale, les valeurs de x sont des écarts types, la mesure habituelle de la variabilité. Pour que la courbe logistique lui soit comparable, on a fixé l'échelle de l'axe horizontal à *l'unité*[67]. Dans le cas des deux courbes, l'axe vertical, $F(x)$, décrit la probabilité d'obtenir une valeur inférieure à x ; c'est-à-dire que $F(x)$ est une fonction intégrant la surface sous la courbe jusqu'au point correspondant sur l'axe des abscisses.

67. Dans le cas de la courbe normale, l'échelle s'exprime en écarts types. Ainsi, si on manipule des poids de poissons, l'axe des x serait en poids de poissons centrés réduits, sans unités. L'échelle de la distribution logistique n'est pas l'écart type, pour des motifs statistiques plutôt complexes. La fixation de l'échelle logistique à l'unité la rend comparable à l'échelle de la courbe normale. Les statisticiens ont l'habitude d'appeler *quantiles* les unités de l'axe et ils légenderaient l'axe avec cette étiquette.

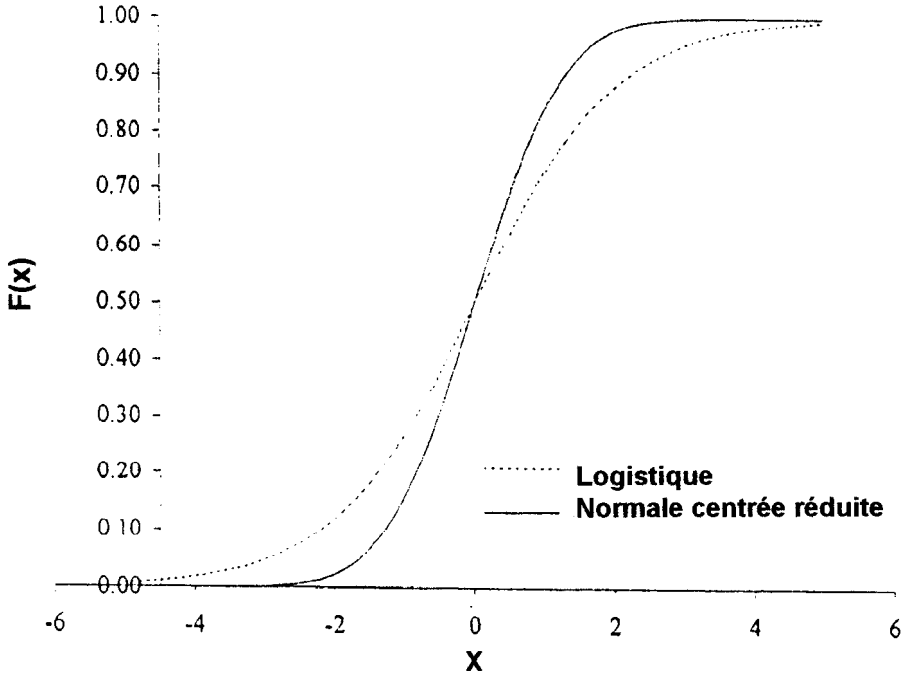


Figure J.1. — Comparaison des distributions logistique et normale. Les distributions sont cumulées, comme on le fait pour les résultats d'un essai de toxicité quantique.

Grâce à la fig. 9, où l'on compare probits et logits, on peut se faire une idée de la relation entre les probits et la courbe dose-effet et de la façon dont on redresse cette courbe.

Pour obtenir les logits, on emploie une transformation assez simple. À une concentration donnée, on divise la proportion d'organismes touchés (p) par ($1 - p$). On trouve le logarithme du quotient, et ce logarithme est le logit que l'on peut utiliser dans l'ajustement d'une régression et l'estimation d'un paramètre de toxicité. La régression est linéaire et l'équation devient :

$$\text{logit}\ (p) = \alpha + \beta X$$

Ainsi, avec les données quantiques telles que les résultats d'un essai de toxicité létale, la transformation a abouti à une relation semblable à la formule bien connue d'une droite (*régression linéaire simple*) : $Y = \alpha + \beta X$. C'est, bien sûr, la relation entre un effet (Y, la *variable dépendante*) et X, la *variable indépendante* (le logarithme de la concentration), expliquée de façon plus approfondie dans les § 6.5.1 et 6.5.2. Cette formule bien connue et l'équation analogue des logits représentent une *régression* de deux paramètres seulement, **α,** l'ordonnée à l'origine, et **β,** la pente de la droite.

Tant dans la régression logistique que probit, pour cet exemple avec des données quantiques, les paramètres **α** et **β** ne peuvent pas être facilement estimés parce que l'on ne peut pas formuler d'équations pour résoudre un paramètre qui ne contient pas l'autre paramètre. La solution est habituellement obtenue par *itération* (§ 4.5.3). Dans le cas de la régression logistique, nous pouvons affirmer, en généralisant, qu'un programme informatique « devine » la valeur du second paramètre, résout l'équation pour le premier paramètre, puis utilise cette estimation pour résoudre le second paramètre. Le processus est répété, en commençant par la valeur que l'on vient d'estimer, du second paramètre, jusqu'à l'arrêt des calculs en vertu de critères prédéterminés qui indiquent l'obtention d'une solution satisfaisante.

La régression ayant été établie pour cette donnée quantique, on peut l'utiliser pour estimer la CE p et ses limites de confiance.

La transformation en logits donne aussi un modèle précieux pour les données quantitatives sur la toxicité sublétale telles que les données sur la croissance et la reproduction. Elle a désormais été adoptée par Environnement Canada comme option pour les analyses des résultats de tels essais de toxicité sublétale (v. le § 6.5.4).

La méthode de Spearman-Kärber

La méthode de Spearman-Kärber (la S.-K.) d'estimation d'une CE 50 a été largement utilisée, particulièrement après avoir été préconisée à l'USEPA. Dans le § 4.5.6, nous avons exposé les exigences de la méthode et la démarche générale à laquelle elle donne lieu. Ici, nous donnerons des détails supplémentaires sur la mécanique interne de la méthode, de sorte que les chercheurs pourront avoir une idée de la façon dont le programme traite les données.

Dans les premières méthodes publiées par Environnement Canada, la S.-K. n'avait pas été recommandée pour l'estimation des CE p, par crainte de l'obtention de résultats divergents par les expérimentateurs qui connaissaient mal les conséquences de l'équeutage des données exprimant la relation dose-réponse (EC, 1992b). En général, on estimait que la S.-K. permettrait de manipuler les données expérimentales par des moyens hors de portée de l'expérimentateur et on craignait que le lissage de données irrégulières ne masque les situations qui méritaient d'être reconnues comme peu communes. Le fameux statisticien Finney a également mis en doute la méthode, parce qu'il est « arithmétiquement possible de l'utiliser dans des situations où sa validité est gravement mise en doute » (Finney, département de la statistique, U. d'Édimbourg, communication personnelle, 1983). En effet, on peut obtenir des résultats anormaux avec des ensembles irréguliers de données (§ 4.4).

Récemment, Environnement Canada a recommandé la version sans équeutage de la méthode S.-K. d'analyse des résultats d'essais révélant un effet partiel, qui ne se prêtent pas à la régression probit ou logit (EC, 2001a, 2004a). Nous préconisons une approche moins rigoureuse. On emploiera la méthode pour les ensembles de données ne comportant qu'un effet partiel, on effectuera les analyses sans équeutage et avec équeutage minimal, et on choisira le paramètre de toxicité qui s'ensuivra et qui sera le plus convenable, à partir d'un tracé des données brutes et des données elles-mêmes.

K.1 Exemples simples de calculs

La méthode de S.-K. repose sur un processus de calcul de la moyenne, essentiellement la moyenne d'un histogramme. La moyenne est assimilée à la médiane, ce qui est vrai pour les distributions symétriques.

Le tableau K.1 offre un exemple *très* simplifié d'un essai avec des poissons pour montrer comment fonctionne la méthode. Les données concernent deux concentrations, 10 et 20 mg/L, avec effet nul chez les 10 poissons à la concentration inférieure et effet total à la concentration supérieure.

À l'aide de la méthode de S.-K., on estime la CE 50 à 15 m/L. Pour expliquer cela de façon anthropomorphique, la concentration inférieure n'est parvenue à tuer aucun poisson, mais la concentration supérieure a suffi pour les tuer tous. Essentiellement, la méthode pose que s'il y avait eu plusieurs concentrations intermédiaires, ils auraient tué les poissons en proportions régulièrement croissantes allant de 0 à 100 %. La méthode attribue donc la moitié de la mortalité au point médian entre les deux concentrations effectivement utilisées (demi-somme des concentrations). Une autre façon, simpliste, d'interpréter cet exemple, c'est, pour le programme, de supposer que les poissons les plus faibles ne seraient pas touchés à 10 mg/L, mais qu'ils le seraient à 11 mg/L, les poissons un petit peu moins faibles ne seraient pas touchés à 11 mg/L, mais le seraient à 12 mg/L, etc. Ainsi, un effet agissant sur le 5^{e} poisson (médian) correspondrait, selon les prévisions, à 15 mg/L, concentration que l'on a adoptée comme la CE 50.

En réalité, le programme utiliserait les logarithmes de la concentration ; dans l'exemple, nous avons utilisé les valeurs arithmétiques dans un souci de simplicité. Il est habituel d'utiliser les logarithmes népériens ($\log_e$) avec la S.-K., mais la base des logarithmes n'a pas d'importance, pour autant que l'on utilise toujours la même.

Tableau K.1. — Exemple simplifié visant à montrer les calculs effectués par la méthode de Spearman-Kärber. Les valeurs arithmétiques des concentrations visent à faciliter la compréhension.

(1)	Concentration	10 mg/L		20 mg/L
(2)	Demi-somme des concentrations, $(C_1 + C_2) \div 2$		15 mg/L	
(3)	Proportion de poissons touchés	0,0		1,0
(4)	Proportion de poissons mourant dans cet intervalle de concentrations (1,0 - 0,0 = 1,0)		1,0	
(5)	Produit des lignes (2) et (4)		15 mg/L	
(6)	CE 50 = somme de tous les éléments de la ligne (5)		15 mg/L	

Habituellement, il y aurait plus de concentrations, comme dans l'exemple plus réaliste du tableau K.2. Cet exemple possède des effectifs exceptionnellement nombreux d'organismes en expérience (poissons en l'occurrence), et les proportions touchées représentent 0 poisson sur 40 (0/40), 1/40, 1/40, 6/38 et 40/40. L'exemple passe exactement par les mêmes étapes que celles du tableau K.1, sauf que l'on a plus de concentrations à manipuler et que l'on utilise les logarithmes népériens des concentrations. L'explication anthropomorphique donnée ci-dessus cesse également de s'appliquer, puisque, à l'étape 4, correspondent quatre proportions de l'effet total. Chacune de ces proportions contribue à l'estimation finale de la CE 50, bien que, dans ce cas, la plus grande contribution provienne de la proportion la plus à droite du tableau. Il importe de conserver beaucoup de chiffres dans les calculs.

Tableau K.2. — Exemple typique de calculs par la méthode de Spearman-Kärber.

(1)	Concentration (mg/L)	15,54		20,47		27,92		35,98		55,52
	logarithme népérien de la concentration	2,7434		3,0190		3,3293		3,5830		4,0167
(2)	Demi-somme		2,8812		3,1742		3,4562		3,7999	
(3)	Proportion touchée	0,0		0,025		0,025		0,158		1,00
(4)	Proportion dans cet intervalle		0,025		0,0		0,133		0,842	
(5)	Produit des lignes 2 et 4		0,07203		0,0		0,45967		3,19952	
(6)	Total des éléments de la ligne 5									3,7312

La CE 50 estimée est de 3,7312, et son antilogarithme est de 41,7 mg/L. Les limites de confiance sont calculées à l'aide de la variance et sont de 39,9 et de 43,7 mg/L. La régression probit donne, dans ce cas, des résultats très semblables.

K.2 Observations sur le mode opératoire

Le lissage des données est une manipulation utilisée dans les calculs de la S.-K. pour obtenir des données monotones. Le lissage peut être nécessaire parce que la méthode exige que l'effet de toute concentration donnée *doit* être supérieur ou égal à l'effet observé à la concentration immédiatement inférieure. Sinon, on prend l'effet moyen de ces deux concentrations, on l'attribue aux deux concentrations et on s'en sert dans les calculs. Cette technique s'appelle

« proportion ajustée attribuée ». Dans le tableau K.2, les deux valeurs de 0,025 énumérées pour la proportion attribuée avaient été ajustées à partir de 0,05, dans la 2e concentration, et de 0,0, dans la 3e.

L'équeutage des extrémités de la distribution est une option des programmes informatiques pour l'application de la S.-K. (« la méthode de Spearman-Kärber avec équeutage [suppression des résultats aberrants]). On peut équeuter mathématiquement 10, 20 % ou même plus des données situées aux extrémités de la courbe cumulée des effets, là où il pourrait y avoir des irrégularités, puis travailler avec la partie centrale de la distribution. Pour l'exemple du tableau K.2, la CE 50 estimée après équeutage de 10 % des résultats serait de serait 42,8 au lieu de 41,7 mg/L ; cela est probablement une meilleure estimation avec un intervalle de confiance plus étroit. Des programmes informatiques (TOXSTAT, CETIS) choisissent automatiquement, sans que l'expérimentateur ait un mot à dire, le taux minimal convenable d'équeutage, qui est considéré comme satisfaisant et que nous recommandons.

On a contesté la validité de l'équeutage. Dans sa version originelle, la S.-K. exigeait des effets de 0 et de 100 % aux extrémités de la distribution. Si l'un ou l'autre des résultats manque et que les deux extrémités de la distribution sont équeutées pour obtenir un ensemble égal de données, **le programme élargit ensuite mathématiquement la distribution à 0 % et 100 %** puis il estime la CE 50. L'équeutage n'est d'aucune aide si l'irrégularité se trouve dans la partie centrale de la distribution. Si de telles irrégularités existaient, il incomberait à l'expérimentateur de les reconnaître et d'y voir comme il se doit.

K.3 Formules mathématiques sous-jacentes à l'analyse de Spearman-Kärber

Les formules utilisées dans la méthode de Spearman-Kärber sont montrées avec deux exemples. Le tableau K.3 montre les calculs se rapportant à l'exemple A du tableau 2 du corps du texte. Le tableau K.4 montre un autre exemple dans lequel on a employé le lissage. La comparaison montre une caractéristique importante des analyses de Spearman-Kärber, c'est-à-dire que le lissage tend à élargir les intervalles de confiance.

On estime le logarithme de la CE 50 à l'aide de l'équation suivante :

$$u = \tfrac{1}{2} \sum_{i=1}^{k-1} \left(p_{i+1} - p_i\right)\left(x_i + x_{i+1}\right) \qquad (K.1)$$

Où :

p_i est la proportion d'organismes (sur n_i organismes) mourant à la i-ième concentration ;
x_i est le logarithme de la i-ième concentration ;
k est le nombre de concentrations ;
p_1 est la mortalité de 0 % ;
p_k est la mortalité de 100 %.

La variance de u est donnée par l'équation suivante :

$$V(u) = \sum_{i=2}^{k-1} \frac{p_i\left(1 - p_i\right)\left(x_{i+1} + x_{i-1}\right)^2}{4\left(n_i - 1\right)} \qquad (K.2)$$

Les intervalles de confiance sont estimés à deux fois l'écart type, soit CE 50 ± 2 fois l'écart type, ce qui suppose que la CE 50 estimée est distribuée comme une variable aléatoire normale.

Tableau K.3. — Calculs appliqués à l'exemple A du tableau 2 selon la méthode de Spearman-Kärber.

Conc. (mg/l)	log de la conc. (x_i)	i	n	Nbre de morts	Prop.[tion] de morts (p_0)	Contrib. à la CE 50 $(p_{i+1} - p_i)(x_i + x_{i+1})$	Contrib. à la variance (équation K.2)
1,8	0,255 273	1	10	0	0	0	
3,2	0,505 15	2	10	2	0,2	0,076 04	0,001 080
5,6	0,748 188	3	10	4	0,4	0,125 3	0,001 633
10	1	4	10	9	0,9	0,437 0	0,000 643 0
18	1,255 273	5	10	10	1	0,112 8	
Sommes :						log (CE 50) = 0,7512	variance du log (CE 50) = 0,003 356

L'intervalle approximatif de confiance au seuil de 95 % de log (CE 50) est ± 2 [racine carrée de la variance du log (CE 50)], soit 0,7512 ± 2 [racine carrée de 0,003 356], dont les limites estimées sont de 0,6353 et de 0,8670. On peut élever ces valeurs à une puissance pour obtenir la CE 50 = 5,64 avec un intervalle de confiance au seuil de 95 % de 4,32 à 7,36. Ce sont essentiellement les valeurs montrées dans le tableau 2.

Si l'effet n'est pas monotone, il faut l'ajuster (le lisser) avant d'employer la méthode de Spearman-Kärber. Les effets adjacents sont combinés conformément à l'équation K.3, qui est adaptée à l'exemple du tableau K.4.

$$p_{3,5} = \frac{e_3 + e_4}{n_3 + n_4} \qquad (K.3)$$

Les données du tableau K.4 peuvent être qualifiées de cas général. Soit la série de concentrations c_1, c_2, c_3, c_4 et c_5. Soit e le nombre de sujets touchés et n le nombre soumis à l'expérience, les effets proportionnels observés sont $p_1 = e_1/n_1$, $p_2 = e_2/n_2$, $p_3 = e_3/n_3$, $p_4 = e_4/n_4$ et $p_5 = e_5/n_5$. Dans cet exemple, $p_3 > p_1$, p_2 et p_4, tandis que $p_4 > p_1$ et p_2. Il faut combiner p_3 et p_4 pour obtenir $p_{3,5}$, comme dans l'équation K.3.

Comme $p_2 < p_{3,5} < p_5$, les calculs peuvent passer à l'estimation du paramètre de toxicité. Si on n'avait pas obtenu la monotonie, on aurait répété le lissage de la même manière.

La CE 50 et son intervalle de confiance au seuil de 95 % sont estimés comme dans le tableau K.3. La CE 50 vaut 5,66 avec un intervalle de confiance au seuil de 95 % de 4,12 à 7,78.

Les effets sont semblables dans ces deux derniers exemples, et les CE 50 sont à peu près égales (5,64 et 5,66). L'intervalle de confiance est quelque peu plus large dans le second cas, pour lequel on a utilisé le lissage (4,12 à 7,78), que dans le cas précédent (4,32 à 7,36). C'est une conséquence typique de la monotonisation.

Tableau K.4. — Calculs selon la méthode de Spearman-Kärber pour les données exigeant lissage.

Conc. (mg/l)	log de la conc. (x_i)	i	n	Nbre de morts	Prop.[tion] de morts (p_0)	Prop. ajustée	Contrib. à la CE 50 ($p_{i+1} - p_i$) ($x_i + x_{i+1}$)	Contrib. à la variance (équation K.2)
1,8	0,255 273	1	10	0	0	0	0,114 063	
3,2	0,50515	2	10	3	0,3	0,3	0,188 001	0,001 417
5,6	0,748 188	3	10	7	0,7	0,6	0	0,001 633
10	1	4	10	5	0,5	0,6	0,451 055	0,001 714
18	1,255 273	5	10	10	1	1		
Sommes :							log (CE 50) = 0,753 119	variance du log (CE 50) = 0,004 764

Renseignements de base sur d'autres méthodes applicables aux données quantiques

L.1 Les méthodes graphiques de Litchfield et Wilcoxon

Cette ancienne méthode « de raccourci » (Litchfield et Wilcoxon, 1949) est désormais une curiosité, mais elle était souvent utilisée jusqu'aux années 1960, avant l'accès facile aux calculateurs électroniques ou aux ordinateurs. Cette méthode se fonde sur une ligne ajustée à vue d'œil, mais elle donne des résultats acceptables. Elle permet d'estimer la CE 50 et ses limites de confiance à 95 %, la pente de la droite ajustée et le khi-deux, comme moyen d'évaluation de l'ajustement.

Si démodée que soit la méthode, nous en décrivons brièvement le mode opératoire au profit de l'expérimentateur. La méthode permettrait d'évaluer les travaux antérieurs et elle reste utile pour vérifier les résultats douteux des programmes informatiques. (En tout cas, la première partie de la méthode Litchfield et Wilcoxon est une droite des probits ajustée à vue d'œil, que l'on recommande dans toutes les analyses visant à déterminer une CE p comme moyen de vérifier les estimations faites par ordinateur.) Il est instructif d'essayer certaines de ces analyses à la main, notamment pour voir comment la pente choisie pour une droite des probits influe sur la largeur de l'intervalle de confiance de part et d'autre de la CE 50. Pour l'ajustement, on peut essayer diverses droites.

Les marches à suivre ont été conçues pour éviter les calculs fastidieux à la main des régressions probit. La pente de la droite ajustée à vue d'œil et son ajustement (khi-deux) sont calculés d'après les écarts par rapport aux points observés sur la ligne. Les limites de confiance au seuil de 95 % par rapport à la CE 50 sont déterminées par l'emploi de nomogrammes, c'est-à-dire des solutions préalablement calculées d'opérations complexes, représentées par trois échelles linéaires imprimées sur une page de façon parallèle. On couche convenablement une règle transparente sur deux échelles linéaires représentant des variables connues puis on lit la réponse (la variable inconnue) sur la 3^{e} échelle que traverse la règle.

Newman fournit une description moderne de la méthode de Litchfield-Wilcoxon (1995) et il remplace les nomogrammes par des calculs arithmétiques. Les calculs sont maintenant assez faciles sur les calculateurs, et on devrait employer des procédures arithmétiques en remplacement des nomogrammes de Litchfield et Wilcoxon.

L.2 Interpolation linéaire

Dans le § 4.5.9, nous avons fait remarquer que l'« interpolation linéaire » a été conçue par l'USEPA comme technique particulière pour les données quantiques, mais qu'elle ne trouve aucune utilisation particulière pour les essais d'Environnement Canada. Si un essai ne montre aucun effet partiel, l'expérimentateur peut utiliser la méthode binomiale qui est l'équivalent exact de l'interpolation linéaire. Pour d'autres configurations de données, on devrait employer des méthodes plus appropriées, recommandées dans le § 4.3. Nous décrivons les méthodes d'« interpolation linéaire » de l'USEPA parce qu'il en est souvent question dans les publications et pour expliquer pourquoi elles ne sont plus exigées au Canada.

Les premiers programmes informatiques d'interpolation linéaire étaient fondés sur l'emploi de valeurs arithmétiques de la concentration (§ 4.5.9), défaut auquel on a remédié dans les méthodes plus récentes d'essai sur les déblais de dragage, qui utilisent des logarithmes (USEPA et USACE, 1994).

La méthode d'interpolation linéaire effectue simplement une interpolation entre deux points et elle ignore les autres éléments de la distribution de l'effet. Si deux concentrations successives ont respectivement produit un effet de 0 % et de 100 %, les calculs pour l'interpolation linéaire pourraient employer l'équation 3 (§ 4.5.7), la formule de la moyenne géométrique.

Une équation plus générale d'interpolation linéaire s'applique aux résultats qui présentent un effet partiel à une ou à plusieurs concentrations. Cela pourrait en théorie être utile dans une situation inhabituelle, bien que nous recommandions d'autres méthodes (§ 4.3). L'équation L.1 provient de l'USEPA et de l'USACE (1994). Avec cette formule, on ne peut pas obtenir de limites de confiance.

$$CE\ 50 = anti\log\frac{(50 - M_I)(\log C_S) + (M_S - 50)(\log C_I)}{M_S - M_I} \qquad (L.1)$$

Où :

C_I = la valeur arithmétique de la concentration qui exerce un effet le plus rapproché de 50 % tout en lui étant inférieur (soit la concentration dite inférieure) ;
C_S = la valeur arithmétique de la concentration qui exerce un effet le plus rapproché de 50 % tout en lui étant supérieur (soit la concentration dite supérieure) ;
M_I = le pourcentage d'effet correspondant à C_I ;
M_S = le pourcentage d'effet correspondant à C_S.

Méthodes non linéaires et méthodes du noyau applicables aux données quantiques

M.1 Régression non linéaire

Kerr et Meador (1996) signalent l'existence de techniques non linéaires pour l'estimation d'une CE p. L'analyse classique opère une transformation en relation linéaire au moyen de probits (ou de logits), et leur exemple employant un modèle linéaire généralisé (GLIM) « utilise le caractère sigmoïde inhérent de la réponse toxicologique ». Il n'est pas clair si l'avantage de ne pas avoir besoin d'une transformation serait annulé par l'inconvénient d'avoir besoin d'un plus grand nombre de paramètres dans l'équation ajustée à la relation. Cependant, leur modèle possède cette qualité de tenir compte de la taille de l'échantillon et, également, d'utiliser les effets à 0 et à 100 % sans nécessiter de facteurs de correction. Le modèle peut estimer la CE p et ses limites de confiance pour toute valeur de *p,* petite ou grande. Cette GLM utilise un « algorithme des moindres carrés itérativement repondérés pour trouver les estimations des paramètres qui réduisent la déviance au minimum ». Kerr et Meador déclarent que les bibliothèques d'analyse des logiciels SAS, Systat et autres possèdent des algorithmes ou des programmes particuliers pour les GLM et qu'elles peuvent servir à l'estimation d'une CL p. Il faut un certain degré de connaissances en statistique pour utiliser la technique à partir de ces bibliothèques.

Malheureusement, Kerr et Meador suivent la même piste naïve que d'autres pour leur analyse, en abandonnant la distribution presque géométrique des concentrations expérimentales dans les données servant d'exemples. L'estimation du paramètre de toxicité pourrait être bien exacte, parce que le modèle peut s'adapter à diverses courbes et qu'il ne dépend pas d'une relation linéaire. Cependant, l'abandon de l'hypothèse géométrique ou logarithmique initiale n'était pas un parti scientifique approprié, et l'utilisation de cette base géométrique pour les concentrations pourrait avoir permis un ajustement plus parcimonieux en paramètres, ce qui constitue un avantage statistique distinct. Cette faute scientifique pourrait être facilement corrigée dans le modèle pour en faire une méthode usuelle.

M.2 Méthodes du noyau

Un *estimateur à noyau* est une fonction de lissage qui régularise une courbe grâce à l'application d'une procédure de moyennage aux points situés à proximité de tout point donné. Le lissage est appliqué à son tour, à chacun des points originellement observés, afin de produire une courbe lisse. On estimerait une CE 50 au 50^{e} centile de la courbe lissée, puis on l'associerait au logarithme de la concentration correspondante.

Pour le lissage, on emploie un procédé de pondération. Pour tout point donné, on accorderait le plus de poids aux observations les plus rapprochées, tandis qu'aux observations plus éloignées on accorderait un poids moindre. Il existe plusieurs techniques de pondération, et les plus intéressants sont les suivantes :

- Le *noyau rectangulaire,* dans lequel on attribue aux points à proximité du point cible un poids unitaire, tandis qu'on attribue à tous les autres points un poids nul (c'est-à-dire que leur contribution est nulle).

- Le *noyau triangulaire,* dans lequel on attribue un poids nul aux points observés plus éloignés du point cible d'une distance spécifiée, tandis que l'on attribue aux observations plus rapprochées un poids allant de 0 à 1.

- Le *noyau gaussien,* dans lequel les poids obéissent à une fonction de densité de probabilité gaussienne ou normale. Cela a pour conséquence d'inclure toutes les observations dans l'estimation de l'observation cible.

L'analyste peut choisir une *fenêtre* pour réguler les poids susmentionnés. Le choix de cette fenêtre influe davantage sur la courbe lissée résultante que le choix de la fonction à noyau (Hastie et Tibshirani, 1990). Ces méthodes, y compris celle de la sélection de la fenêtre optimale, sont discutées par Härdle (1991) et Scott (1992).

Les méthodes du noyau sont avantageuses pour les essais de toxicité, puisqu'elles ne sont pas paramétriques et qu'on pourrait les appliquer quand il n'y a pas d'effet partiel dans l'ensemble de données. Des méthodes potentielles n'ont pas encore été évaluées pour ce qui concerne les types des données qui pourraient résulter d'essais d'écotoxicité au Canada, mais certaines évaluations de leur pertinence ont été faites (Kappenman, 1987). Müller et Schmidt (1988) ont évalué de très grands ensembles simulés de données (48 concentrations avec 48 organismes par concentration). Si les données étaient non sigmoïdes, l'analyse par la méthode du noyau a permis d'estimer une CE 50 dont la variance était plus petite de 40 à 70 % que celle que l'on aurait obtenue par régression probit, résultat très impressionnant. Cependant, les données sigmoïdes seraient plus habituelles dans les résultats des essais, et, dans leur cas, la variance était plus grande de 20 à 30 % que celle que l'on aurait obtenue par régression probit.

Annexe N

Estimations ponctuelles applicables aux données quantitatives par lissage et interpolation

N.1 Préparatifs pour l'analyse

Nous exposons les étapes de la méthode de lissage et d'interpolation avec beaucoup plus de détails que dans le § 6.4.2. On peut calculer la CI p à la main, si on le désire (l'explication suit). L'exemple est celui du poids des poissons à la fin de l'essai.

(1) Calculer le poids moyen des poissons détenus dans chaque répétition de chaque concentration (y compris des poissons témoins). À partir des valeurs des répétitions, calculer le poids moyen total à chaque concentration.

(2) Tracer le graphique des poids moyens en fonction du logarithme des concentrations (sur l'axe horizontal). Par ce moyen, on vérifie subjectivement la qualité des données.

(3) Si nécessaire, lisser les données. Aucun lissage n'est nécessaire si le poids moyen global reste le même ou diminue à chaque pas d'augmentation de la concentration, en partant du témoin jusqu'à la concentration maximale. Si, à une concentration quelconque, cette condition n'est pas respectée, il faut lisser la courbe. Le processus doit utiliser la moyenne pondérée des moyennes (voir le texte qui suit).

- Si le poids moyen à la concentration minimale est supérieur au poids du témoin, calculer la demi-somme de ces deux poids moyens et l'utiliser pour le témoin et pour la concentration minimale.

- Si le poids moyen correspondant à la 2^{e} concentration minimale est supérieur au poids moyen correspondant à la concentration minimale, calculer la demi-somme de ces deux poids moyens et l'utiliser pour les deux concentrations minimales. Répéter cette étape pour chaque paire de concentrations croissantes, jusqu'à la concentration maximale.

- Si les nouveaux poids moyens ne sont pas égaux ou ne diminuent pas monotonement, répéter le lissage de la paire ou des paires appropriées de concentrations, en pondérant chaque valeur utilisée dans le calcul des moyennes en fonction du nombre de concentrations que cette valeur représente à l'origine[68].

- Répéter les deux étapes qui précèdent tant que l'ensemble de résultats n'est pas monotone.

68. Le lissage se fait d'une façon particulière. Si la concentration minimale a donné un poids moyen de, disons, 14 unités, soit davantage que le poids des témoins, de 8 unités, on calculerait la demi-somme de ces poids dans le premier cycle de lissage. Le résultat (11 unités) représente l'effet chez le témoin et à la concentration minimale. On passe ensuite aux deux concentrations supérieures à la concentration minimale et ainsi de suite, à toutes les concentrations deux à deux. Le deuxième cycle de lissage débuterait de nouveau avec le témoin ; si à la deuxième concentration minimale correspondait la moyenne de 13, ce qui est davantage que la nouvelle valeur calculée pour la concentration minimale (et le témoin), on calculerait la demi-somme de 13 et de 11, *et on utiliserait le résultat pour le témoin et les deux concentrations minimales.* La nouvelle moyenne serait pondérée en fonction du nombre d'observations originelles, dans ce cas la valeur de 11 aurait deux fois le poids de 13. Si, pour chaque concentration, on disposait de 4 observations (répétitions), le calcul serait le suivant : $[(8 \times 11) + (4 \times 13)] / 12 = 11{,}7$. Sinon, on pourrait revenir aux observations originelles et en calculer la moyenne : $[(4 \times 8) + (4 \times 14) + (4 \times 13)] / 12 = 11{,}7$. La valeur 11,7 représente désormais l'effet observé chez le témoin et à chacune des deux concentrations minimales. À noter que la deuxième concentration minimale a été incluse dans le lissage parce le poids moyen qui lui correspondait était plus élevé que la *demi-somme* correspondant au témoin et à la concentration minimale ; le poids qui lui correspondait n'était pas effectivement supérieur au poids qui correspondait originellement à la concentration minimale. C'est pourquoi, si l'on effectue le processus à la main, il serait mieux de lisser les paires initiales de valeurs, puis de répéter le cycle.

- Les nouvelles moyennes servent de données d'entrée pour l'analyse. Dans l'équation N.1 (§ N.2), *M* symbolise le nouveau poids moyen, M_1 le poids du témoin et M_j une concentration à préciser. Toutes les concentrations originelles subsistent dans l'analyse, peut-être avec un effet modifié (lissé).

- Le lissage peut être une manipulation risquée de l'ensemble de données, particulièrement si cet ensemble est irrégulier ou hormétique. Nous décrivons des problèmes potentiels dans le § 6.4.1. Il importe de déterminer si le paramètre calculé de toxicité est acceptable, lorsqu'on le compare aux données originelles (brutes).

N.2 Estimation d'une concentration inhibitrice (CI p)

La méthode d'estimation semble complexe, lorsque l'on la décrit étape par étape, mais elle n'est simplement qu'une interpolation linéaire entre les deux concentrations encadrant l'effet recherché. Les étapes exposées ci-dessous conduisent à une formule définissant le même mode opératoire. L'analyse reprend à partir de l'étape 3.

(4) Décider de la valeur de *p*. Soit $p = 25$, de sorte que la CI 25 représentera la concentration correspondant à un poids inférieur de 25 % à celui des poissons témoins.

(5) Examiner les données pour déterminer les deux concentrations qui encadrent une réduction de 25 % du poids. À partir d'ici, nous n'utilisons que ces deux concentrations et les poids moyens qui leur correspondent.

(6) Calculer le poids représentant le paramètre de toxicité. C'est 75 % du poids des poissons témoins, c'est-à-dire multiplier M_1 par 0,75.

(7) Du produit de l'étape 6, soustraire le poids (M_j) correspondant à la concentration immédiatement inférieure à la CI 25. Le résultat sera négatif.

(8) Du poids moyen correspondant à la concentration immédiatement supérieure à la CI 25, soustraire le poids (M_j) correspondant à la concentration immédiatement inférieure à la CI 25. Normalement, cette différence (appelée M_{diff}) est négative.

(9) Diviser le résultat de l'étape 7 par celui de l'étape 8.

(10) Calculer la différence entre le logarithme de la concentration immédiatement inférieure (C_j) et le logarithme de la concentration immédiatement supérieure à la CI 25 (il importe de soustraire le logarithme de la concentration inférieure de celui de la concentration supérieure). On appelle le résultat C_{diff}.

(11) Multiplier le résultat de l'étape 9 par celui de l'étape 10. Cela représente l'augmentation de concentration de la CI 25 par rapport à la concentration (C_j) qui lui est immédiatement inférieure.

(12) Ajouter le résultat de l'étape 11 à la concentration logarithmique (C_j) immédiatement inférieure à la CI 25. Le résultat est la CI 25 sous forme logarithmique.

$$CI\ p = C_j + \frac{0{,}75\ M_1 - M_j}{M_{diff}} \times C_{diff} \qquad (N.1)$$

Où :

C_j = le logarithme de la concentration immédiatement inférieure à la CI 25.

C_{diff} = la différence entre les logarithmes des concentrations adjacentes à la CI 25, la supérieure moins l'inférieure.

M_1 = l'effet moyen (poids des poissons) chez le groupe témoin.
M_j = l'effet moyen correspondant à la concentration immédiatement inférieure à la CI 25.
M_{diff} = la différence entre l'effet moyen à la concentration supérieure et l'effet à la concentration inférieure (le signe est important).

Si aucune concentration expérimentale n'est inférieure et supérieure à la CI p, il est impossible d'estimer cette dernière. On peut seulement affirmer que la CI p est inférieure à la concentration expérimentale minimale ou supérieure à la concentration maximale, selon le cas.

N.3 Limites de confiance et le programme informatique ICPIN

Il faut un ordinateur pour appliquer la technique « bootstrap » pour la détermination des limites de confiance à 95 % de la CI p. Cela nécessite le calcul d'une série de CI p qui pourraient avoir été obtenues, d'après des rééchantillonnages des observations originelles (répétitions). À partir de la série de CI p hypothétiques, il est possible de calculer des limites acceptables de confiance pour la CI p estimée [69].

Le programme ICPIN, offert dans des progiciels commerciaux, tourne sur ordinateur personnel ; cependant, il est libre de droits, et on peut en obtenir des exemplaires de l'USEPA [70]. Le programme ICPIN est facile à utiliser, ses instructions sont claires, et la manipulation des données se passe d'explications [71]. Il ne faudrait pas utiliser BOOTSTRP, une première version du programme.

Le programme ICPIN effectue toutes les étapes (1 à 11) exposées dans le § N.2, et on y saisit les observations brutes. Pour obtenir un résultat juste, **il faut,** à l'encontre des instructions du programme, **saisir les logarithmes des concentrations d'essai.** À la fin, on peut convertir les valeurs logarithmiques de la CI p et de ses limites de confiance en valeurs arithmétiques pour en faciliter la compréhension. Certains programmes commerciaux fondés sur ICPIN

---

69. Au moins 240 nouvelles estimations de CI p hypothétiques devraient être faites. Chaque estimation découle du rééchantillonnage des données correspondant à chaque concentration expérimentale, qui permet à tout point de donnée d'être choisi plus d'une fois (« rééchantillonnage au hasard avec remise »). Le programme informatique effectue l'échantillonnage au hasard. Par exemple, les données sur l'effet saisies dans le programme pourraient être le poids total (ou moyen) des poissons dans chacune des quatre enceintes (répétitions) correspondant à chaque concentration expérimentale. L'ordinateur choisirait quatre valeurs pour représenter une concentration, parmi les quatre poids connus à cette concentration. Il choisirait chacune des quatre valeurs dans la même gamme des quatre poids (« échantillonnage avec remise »), de sorte que chaque échantillonnage inclurait probablement certains poids deux fois ou plus et pourrait ne pas comprendre un ou plusieurs poids. Une sélection semblable serait faite à chaque concentration expérimentale, puis une CI p hypothétique serait calculée. Ensuite l'ordinateur recommencerait avec un autre jeu de sélections aléatoires parmi les mêmes données, avec un autre calcul de la CI p, ainsi de suite.

Selon les sélections fortuites, on pourrait obtenir des jeux de données et des CI p tout à fait variables. Une variation plus grande dans les données originelles entraîne un plus grand étalement des CI p calculées. La série de 240 (au moins) CI p hypothétiques aura sa propre distribution. Les concentrations qui délimitent 2,5 % des CI p hypothétiques aux deux extrémités de la distribution servent à estimer les limites de confiance de la CI p effectivement obtenue dans l'expérience. La technique « bootstrap » a été proposée par Efron (1982) et discutée par Marcus et Holtzman (1988).

Si les limites étaient estimées à partir de seulement 80 échantillonnages bootstrap, les estimations risqueraient d'être instables (USEPA, 1995). Le premier programme informatique BOOTSTRP tendait à donner un intervalle de confiance étroit, qui péchait par optimisme, et cela avait été noté dans le compte rendu de la réunion du Groupe consultatif canadien sur la statistique (Miller *et al.,* 1993). Cette tendance était particulièrement évidente quand le nombre de répétitions était petit, par ex. deux par concentration.

70. La source du programme est l'EMSL-CINCINNATI, United States Environmental Protection Agency,, 3411 Church Street, Cincinnati (Ohio) 45244, États-Unis. Dans la pratique, puisque le programme est libre de droits, de nombreux chercheurs en ont obtenu copie de confrères d'un autre laboratoire. Comme nous l'avons mentionné, ce programme fait partie intégrante de programmes commerciaux utilisés en toxicologie.

71. La facilité d'emploi n'est pas nécessairement vraie dans le cas des programmes commerciaux qui incorporent le programme ICPIN, comme il est mentionné dans le § N.4.

offrent la possibilité de transformer la concentration (ou la dose) en logarithmes de base 10 ; cependant, **il faudrait s'assurer que la transformation est en réalité conservée et utilisée** dans les calculs (§ 2.3.2 et annexe N).

Le programme ICPIN manipule jusqu'à 12 concentrations, y compris celle du groupe témoin, et jusqu'à 40 éléments par concentration. Ces « éléments » doivent être de véritables répétitions. Par exemple, si on pesait 10 poissons dans un récipient à une concentration donnée, les poids ne seraient pas des répétitions ; le poids total ou le poids moyen serait la valeur à saisir dans ICPIN, en tant qu'une répétition. Il semble peu probable que 40 répétitions soient employées un jour dans des essais exécutés conformément aux méthodes d'Environnement Canada. Aux diverses concentrations, le nombre d'éléments n'a pas besoin d'être égal. Le degré (*p*) d'effet choisi comme paramètre de toxicité peut varier de 1à 99 %.

L'expérimentateur doit préciser le nombre de rééchantillonnages dans la partie du programme appliquant la technique bootstrap. Le nombre peut varier de 80 à 1 000 par pas de 40 ; le nombre habituellement recommandé est d'au moins 240 (Norberg-King, 1993), et il n'y a pas de raison pour ne pas choisir un nombre élevé, disons 800. Si on saisit plus de six données (répétitions) par concentration, le programme calcule les limites « originelles » de confiance au seuil de 95 %. Si on saisit moins de sept données, le programme ICPIN de 1993 (version 2.0) calcule les limites de confiance originelles et « étendues », et l'investigateur devrait utiliser les valeurs étendues, qui résultent d'une tentative de permettre des estimations excessivement optimistes des limites par la technique bootstrap.

Le programme reproduit des tableaux de données et des calculs préliminaires. On devrait utiliser la CI p estimée par interpolation linéaire. Le programme imprime les limites de confiance, originelles ou originelles et étendues, comme nous venons de le mentionner. Il imprime également une valeur moyenne de la CI p, résultant de l'échantillonnage bootstrap, et son écart type ; ce n'est pas le résultat de l'essai de toxicité et il ne faut pas le signaler comme tel.

N.4 Programmes commerciaux comprenant le programme ICPIN

Des progiciels commerciaux renferment des versions d'ICPIN ainsi que d'autres programmes d'analyse des résultats des essais de toxicité. Au moment d'écrire ces lignes, trois progiciels sont communément utilisés : la version 3.5 de TOXSTAT (1996), la version 5.0 de TOXCALC (1994) et CETIS (2001). Ces progiciels ont été utilisés par les expérimentateurs canadiens. Comme les programmes sont modifiés de temps à autre et qu'il en apparaît de nouveaux, nous ne formulerons ici que des observations générales.

Les programmes commerciaux suivent habituellement de près les modes opératoires de l'USEPA et tendent à produire des renseignements visant à satisfaire aux exigences de cet organisme, parfois en employant des formulaires de déclaration de l'administration. Les méthodes d'essai et les rapports ne satisfont pas nécessairement aux exigences d'Environnement Canada. TOXCALC exige la saisie fastidieuse de beaucoup de renseignements accessoires, qui ne sont pas nécessaires.

Les programmes commerciaux tendent à être rédigés pour une application dans les ordinateurs personnels actuellement utilisés. Les progiciels commerciaux n'étaient pas aussi faciles et évidents que le programme ICPIN lui-même, pour ce concerne le paramétrage, la saisie des données et l'analyse. Certaines vieilles versions des programmes commerciaux employaient des méthodes particulières de saisie des données ou étaient récalcitrantes dans leur fonctionnement. Les notices omettaient des sujets ou étaient difficiles à comprendre. Les programmes commerciaux n'offraient pas un accès téléphonique gratuit en cas de besoin. L'expérimentateur devrait exécuter les fichiers à titre d'exemple, le cas échéant, pour se familiariser avec le formatage requis.

Comme nous l'avons mentionné, l'expérimentateur doit s'assurer d'utiliser le logarithme des concentrations dans l'analyse, ce qui exigera probablement la saisie des logarithmes dans la plupart des logiciels. TOXSTAT 3.5 offre la transformation des concentrations en logarithmes, mais pour conserver les logarithmes au cours de l'analyse, il faut choisir cette option et en commander l'exécution avant de procéder à l'analyse.

Estimation des CI p par régression linéaire et non linéaire

Dans le § 6.5.8, nous avons exposé le mode opératoire général à suivre pour la méthode usuelle, à Environnement Canada, de régression des données quantitatives des essais de toxicité. Dans la présente annexe, nous exposons les consignes générales, point par point, pour réaliser une analyse. Les méthodes statistiques sont identiques à celles qui sont exposées comme modes opératoires normalisées dans les méthodes récentes d'Environnement Canada d'estimation de la toxicité d'un sol (2004a, b et 2007).

O.1 Introduction

Dans la présente annexe, nous donnons des conseils sur l'emploi de la régression linéaire et de la régression non linéaire pour l'estimation de CI p, d'après les relations concentration-réponse dans les données quantitatives. Nous reprenons, en l'adaptant, l'approche décrite par Stephenson *et al.* (2000). Les marches à suivre concernent la version 11.0 de SYSTAT [72] ; cependant, on peut utiliser tout logiciel convenable. Ces techniques de régression s'appliquent le mieux aux données continues, obtenues grâce à des plans d'expérience prévoyant au moins 10 concentrations ou traitements, y compris le ou les témoins. Les plans d'expérience de la mesure des effets d'une exposition prolongée sur le ver de terre *Eisenia andrei,* des collemboles (par ex. *Folsomia candida* ou *Onychiurus folsomi*) ou la croissance végétale sont résumés dans le tableau O.1.

Nous présentons dans la figure 16 une vue d'ensemble du processus général utilisé pour évaluer la mesure dans laquelle un ensemble de données se prête à ces régressions.

Nous encourageons le lecteur à consulter, avant d'analyser les données, les passages appropriés du présent guide statistique, de même que les passages appropriés sur les plans d'expérience et les analyses de régression dans les méthodes propres aux vers de terre, aux végétaux et aux collemboles (EC, 2004a, b et 2007). Nous avons répété dans la présente annexe certains conseils donnés dans ces documents.

O.2 Régressions linéaires et non linéaires

O.2.1 Création de tableaux de données

L'analyse statistique doit utiliser les logarithmes des concentrations ($\log_{10}$ ou $\log_e$). Si des concentrations sont inférieures à l'unité (1) [par ex. 0,25], on peut transformer les unités de concentration (par ex. les mg/kg en µg/g) à l'aide d'un multiplicateur (1 000 en l'occurrence) ; on exprime alors les concentrations modifiées en logarithmes. On peut enregistrer les valeurs logarithmiques dans la feuille de calcul électronique d'origine ou effectuer le changement lors du transfert des données originelles dans le fichier de données de SYSTAT. Avant de les publier, on devrait transformer les CI p et leurs limites de confiance en valeurs arithmétiques, afin de les rendre plus intelligibles.

(1) Ouvrir le fichier renfermant l'ensemble de données dans une feuille de calcul électronique.

(2) Ouvrir le programme SYSTAT. Dans la fenêtre principale, cliquer sur **File** (Fichier), **New** (Nouveau) et **Data** (Données). On ouvre ainsi un tableau vide. Il faut insérer le nom des variables dans l'en-tête de la colonne en cliquant deux fois sur le nom d'une variable, ce qui ouvre la fenêtre '**Variable Properties**' (« Propriétés des variables »). Insérer un nom approprié pour la variable recherchée dans la zone '**Variable Name**' (« Nom de la

72. On peut se procurer la version la plus récente (par ex. 11.0) de SYSTAT™ auprès de SYSTAT Software Inc., 501 Canal Blvd, Suite C, Point Richmond, CA 94804-2028, États-Unis ; tél. : 800 797-7401 ; site Web : www.systat.com/products/Systat/.

Tableau O.1. — Sommaire des plans d'expérience des méthodes biologiques d'Environnement Canada pour les essais de toxicité d'un sol pour la croissance de végétaux ou la reproduction de vers de terre et de collemboles.

Variable	Ver de terre	Végétal	Collembole	
Espèce	*Eisenia andrei* ; adultes avec clitellum, dont le poids frais individuel varie de 250 à 600 mg	Diverses espèces	*Folsomia candida* ; âges synchronisés ; 10–12 jours après l'éclosion	*Onychiurus folsomia* ; adultes dont la longueur du corps excède 2 mm ; pas de synchronisation de l'âge ; 5 mâles et 5 femelles
Durée de l'essai	56 jours (8 semaines)	14 ou 21 jours ; selon l'espèce	28 jours	35 jours
Nombre de répétitions	10 par traitement	6 par traitement témoin ; 4 pour chaque concentration inférieure ; 3 pour les concentrations médianes et maximales	Au moins 3 répétitions par traitement ; au moins 5 par traitement témoin	Au moins 10 par traitement, y compris le témoin
Nombre de traitements	Sol témoin négatif et au moins 7 concentrations ; au moins 10 concentrations plus un témoin négatif fortement recommandés	Sol témoin négatif et au moins 9 concentrations expérimentales	Sol témoin négatif et au moins 7 concentrations ; au moins 10 concentrations plus un témoin négatif fortement recommandés	Sol témoin négatif et au moins 7 concentrations ; au moins 10 concentrations plus un témoin négatif fortement recommandés
Paramètres statistiques	**Quantiques :** les méthodes de la présente annexe ne sont pas appropriées. Utiliser des méthodes quantiques s'il existe une relation convenable entre la concentration et l'effet.			
	• Pourcentage moyen de survie des adultes dans chaque traitement, au jour 28 • Calculer la CL 50 28 j (méthodes quantiques)	• Pourcentage moyen de levée à chaque traitement • Calculer la CE 50 14 j ou 21 j par des méthodes quantiques	• Pourcentage moyen de survie des adultes dans chaque traitement, au jour 28 • Calculer la CL 50 28 j (méthodes quantiques)	• Pourcentage moyen de survie des adultes dans chaque traitement, au jour 35 • Calculer la CL 50 35 j (méthodes quantiques)
	Quantitatifs : estimer la CI p (par ex. la CI 50 et/ou la CI 25)			
	• Nombre moyen et masse sèche des jeunes survivant à chaque traitement, au jour 56 • CI p pour la masse sèche et le nombre de jeunes vivants	• Longueur moyenne et masse sèche des pousses et des racines à chaque traitement, au jour 14 ou 21 • CI p pour la longueur et la masse sèche moyenne des pousses et des racines	• Nombre moyen de jeunes survivant à chaque traitement, au jour 28 • CI p pour le nombre de jeunes vivants engendrés	• Nombre moyen de jeunes survivant à chaque traitement, au jour 35 • CI p pour le nombre de jeunes vivants engendrés

variable ») et choisir le type de la variable ; on peut insérer des observations supplémentaires dans la zone '**Comments:**' (« Observations »). Par exemple, on pourrait utiliser les noms suivants de variables :

conc	=	concentration ou traitement ;
logconc	=	valeur de la concentration ou du traitement en $\log_{10}$;
rep	=	répétition à l'intérieur d'un traitement donné ;
juveniles	=	nombre de jeunes engendrés ;
jdrywt	=	poids sec des jeunes engendrés ;
mnlengths	=	longueur moyenne des pousses ;
mnlengthr	=	longueur moyenne des racines ;
drywts	=	masse sèche des pousses ;
drywtr	=	masse sèche des racines.

(3) Transférer les données en copiant et collant chaque colonne de la feuille de calcul renfermant les concentrations, les répétitions et les valeurs moyennes connexes dans le tableau de données de SYSTAT*.

(4) Enregistrer les données en cliquant sur **File** (Fichier), puis sur **Save As** (Enregistrer sous), ce qui ouvre une fenêtre intitulée **'Save As**' (« Enregistrer sous »). Employer le codage approprié pour enregistrer le fichier de données. Sélectionner **Save** (Enregistrer) après avoir saisi le nom du fichier.

(5) Enregistrer le nom du fichier de données de SYSTAT dans la feuille de calcul électronique renfermant les données d'origine.

(6) S'il faut transformer les données (c'est-à-dire les concentrations d'essai) en logarithmes, cliquer sur **Data** (Données), **Transform** (Transformer), puis **Let...** (Soit...). Une fois dans la fonction **Let...** (Soit...), choisir l'en-tête approprié de colonne pour le format souhaité (par ex. logconc), puis choisir **Variable** (Variable) dans la zone '**Add to**' (« Ajouter à ») pour insérer la variable dans la zone intitulée '**Variable:**' (« Variable »). Choisir le code approprié (par ex. L10 pour la transformation en $\log_{10}$ ou LOG pour la transformation en logarithme naturel) dans la zone '**Functions:**' (« Fonctions ») [la zone '**Function Type:**' (« Type de fonction ») devrait être **Mathematical** (Mathématique)], puis cliquer sur **Add** (Ajouter) pour insérer la fonction dans la zone '**Expression:**' (« Expression »). Choisir l'en-tête de colonne renfermant la version arithmétique des données (c'est-à-dire 'conc' pour la concentration ou le traitement), puis **Expression** (Expression) dans la zone '**Add to**' (« Ajouter à ») pour insérer la variable dans la zone '**Expression:**' (« Expression »). S'il faut un facteur de multiplication pour ajuster la concentration avant sa transformation logarithmique, on peut réaliser cette étape dans la zone '**Expression:**' (« Expression ») [par ex. L10 (conc. × 1 000)]. Cliquer sur **OK** quand on a effectué toutes les opérations voulues. Les données logarithmiques apparaîtront dans la colonne appropriée. *Enregistrer les données* (c'est-à-dire cliquer sur **File** [Fichier], puis sur **Save** [Enregistrer]).

On ne peut pas fournir le $\log_{10}$ de la concentration du témoin négatif parce que le $\log_{10}$ de 0 est indéfini. Il faut donc affecter au témoin une valeur très faible (par ex. 0,001) que l'on sait ou que l'on pose être une concentration n'exerçant aucun effet. Cela permettra l'inclusion de ce traitement dans l'analyse et de le différencier des autres concentrations sous forme logarithmique.

(7) À partir du tableau de données, calculer et consigner la moyenne des témoins négatifs pour la variable à l'étude. Chaque paramètre de toxicité est analysé de façon indépendante. La valeur moyenne de ces données témoins sera nécessaire à l'estimation des paramètres du modèle. En outre, déterminer la valeur maximale de l'ensemble de données correspondant à cette variable particulière et l'arrondir au nombre entier supérieur le plus rapproché. Ce nombre sert de valeur maximale à l'axe des *y* (c'est-à-dire « y_{max} ») lors de la création d'un graphique des données soumises à la régression.

O.2.2 Création d'un nuage de points ou d'un graphique linéaire

Les diagrammes de dispersions (nuages de points) et les graphiques linéaires donnent une idée de l'allure de la courbe concentration-réponse correspondant à l'ensemble de données. On peut ensuite comparer la forme de la courbe à chaque modèle (fig. O.1) de façon à retenir le ou les modèles les plus appropriés. On devrait ensuite utiliser chacun des modèles retenus pour analyser les données, puis revoir chaque modèle après l'analyse. On retient le modèle qui présente le meilleur ajustement aux données.

(1) Cliquer successivement sur **Graph** (Graphique), **Summary Charts** (Graphiques sommaires), **Line...** (...Linéaires). Choisir la variable indépendante (par ex. logconc), puis sur **Add** (Ajouter) pour insérer la variable dans la zone '**X-variable(s):**' (« Variable[s] x »). Choisir la variable dépendante en examen, puis cliquer sur **Add** (Ajouter) pour insérer la variable dans la zone '**Y-variable(s):**' (« Variable[s] y »). Cliquer sur **OK**. Un graphique apparaîtra dans l'**Output Pane**' (Sous-fenêtre des résultats) de l'écran principal de SYSTAT renfermant les valeurs moyennes correspondant à chaque traitement. Pour visualiser une version plus grande du graphique, cliquer simplement sur l'onglet '**Graph Editor**' (« Éditeur de graphiques ») situé sous la fenêtre centrale. On peut également visualiser le nuage de points correspondant aux données en cliquant sur **Graph** (Graphique), **Plots** (Tracés), puis **Scatterplot...** (Nuages de points...), puis en suivant les même instructions pour l'insertion des variables *x* et *y*. Les graphiques donneront une idée de l'allure générale de la courbe concentration-réponse, qui permettra de retenir le ou les modèles susceptibles de fournir le meilleur ajustement aux données. Ils montreront aussi la valeur approximative de la CI p à laquelle on s'intéresse.

L'écran principal de SYSTAT est divisé en trois parties. Dans la sous-fenêtre de gauche ('**Output Organizer**' [Organisateur des résultats]) on trouve l'énumération de toutes les fonctions appliquées (par ex. les graphiques) — chaque fonction peut être visualisée simplement par la sélection de l'icône voulue. Le côté droit constitue la fenêtre centrale dans laquelle on peut visualiser la restitution générale de toutes les fonctions appliquées (par ex. la régression, les graphiques). Les onglets sous la fenêtre centrale permettent de commuter entre le fichier de données (utiliser l'onglet '**Data Editor**' [« Éditeur de données »]), les graphiques individuels ('**Graph Editor**' [« Éditeur de graphiques »]) et les résultats ('**Output Pane**' [sous-fenêtre des résultats]). On peut visualiser individuellement les divers graphiques produits dans le '**Graph Editor**' (« Éditeur de graphiques ») en sélectionnant le graphique voulu dans la partie gauche de l'écran (onglet '**Output Organizer**' [« Organisateur des résultats »]). La partie inférieure de la fenêtre affiche les codes de commande utilisés pour obtenir les fonctions voulues (régression et construction de graphiques). L'onglet '**Log**' (« Journal ») de cet écran de commandes permet l'affichage de l'historique de toutes les fonctions ayant été appliquées.

(2) Estimer visuellement et consigner une estimation de la CI p (par ex. la CI 50) pour l'ensemble de données. Par exemple, pour la CI 50, diviser la moyenne des mesures relatives aux témoins par 2 et trouver cette valeur sur l'axe des ordonnées (*y*). Projeter une ligne horizontale partant de cet axe jusqu'à ce qu'elle coupe le nuage de points. Tirer une ligne verticale vers l'axe des abscisses et consigner la concentration ainsi trouvée comme l'estimation approximative de la CI 50.

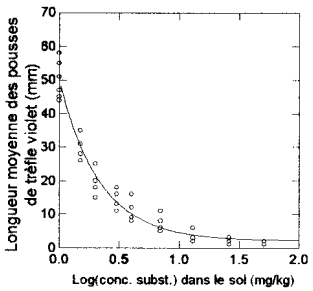
(3) À l'aide des nuages de points ou des graphiques linéaires, retenir le ou les modèles susceptibles de mieux décrire la tendance de la relation concentration-réponse (cf. fig. O.1, pour un exemple de chaque modèle).

O.2.3 Estimation des paramètres du modèle

(1) Cliquer sur **File** (Fichier), **Open** (Ouvrir), puis **Command** (Commandes).

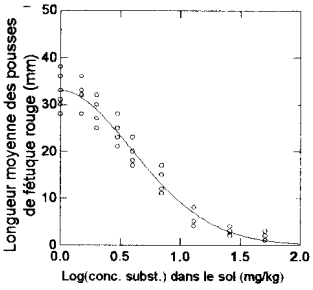
Modèle exponentiel
CI 50 : mnlengths = a*exp(log((a-a*0.5-b*0.5)/a)*(logconc/x))+b
CI 25 : mnlengths = a*exp(log((a-a*0.25-b*0.75)/a)*(logconc/x))+b

Où :
a = l'ordonnée à l'origine (réaction des organismes témoins) ;
x = la CI p pour l'ensemble de données ;
logconc = la valeur logarithmique de la concentration d'exposition ;
b = un paramètre d'échelle (estimé entre 1 et 4).



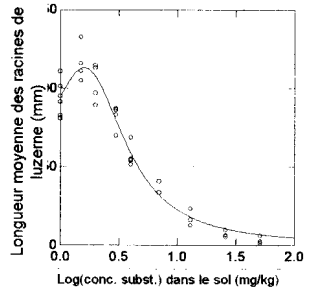
Modèle de Gompertz
CI 50 : mnlengths = g*exp((log(0.5))*(logconc/x)^b)
CI 25 : mnlengths = g*exp((log(0.75))*(logconc/x)^b)

Où :
g = l'ordonnée à l'origine (réaction des organismes témoins) ;
x = la CI p pour l'ensemble de données ;
logconc = la valeur logarithmique de la concentration d'exposition ;
b = un paramètre d'échelle (estimé entre 1 et 4).



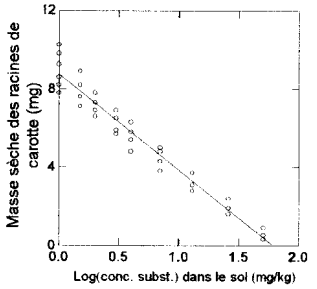
Modèle hormétique
CI 50 : mnlengthr = (t*(1+h*logconc))/(1+((0.5+h*logconc)/0.5)*(logconc/x)^b)
CI 25 : mnlengthr = (t*(1+h*logconc))/(1+((0.25+h*logconc)/0.75)*(logconc/x)^b)

Où :
t = l'ordonnée à l'origine (réaction des organismes témoins) ;
h = l'effet hormétique (estimé entre 0,1 et 1) ;
x = la CI p pour l'ensemble de données ;
logconc = la valeur logarithmique de la concentration d'exposition ;
b = un paramètre d'échelle (estimé entre 1 et 4).



Modèle linéaire
CI 50 : drywtr = ((-b*0.5)/x)*logconc+b
CI 25 : drywtr = ((-b*0.25)/x)*logconc+b

Où :
b = l'ordonnée à l'origine (réaction des organismes témoins) ;
x = la CI p pour l'ensemble de données ;
logconc = la valeur logarithmique de la concentration d'exposition ;



Modèle logistique
CI 50 : drywts = t/(1+(logconc/x)^b)
CI 25 : drywts = t/(1+(0.25/0.75)*(logconc/x)^b)

Où :
t = l'ordonnée à l'origine (réaction des organismes témoins) ;
x = la CI p pour l'ensemble de données ;
logconc = la valeur logarithmique de la concentration d'exposition ;
b = un paramètre d'échelle (estimé entre 1 et 4).

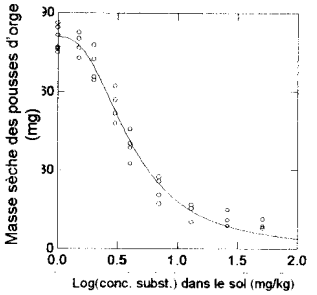


Figure O.1. — Équations d'après la version 11.0 de SYSTAT, pour des modèles de régression linéaire et non linéaire et exemples de graphiques pour chaque modèle.

(2) Ouvrir (ou créer) le fichier renfermant les codes de commande du modèle retenu au § O.2.2 (c'est-à-dire choisir le fichier approprié, puis cliquer sur **Open** [Ouvrir]):

nonline.syc	=	modèle exponentiel ;
nonling.syc	=	modèle de Gompertz ;
nonlinh.syc	=	modèle hormético-logistique ;
linear.syc	=	modèle linéaire ;
nonlinl.syc	=	modèle logistique.

Le fichier fournit les codes de commande du modèle choisi en vertu de l'onglet approprié de la zone de l'éditeur de commandes au bas de l'écran principal. Tous les codes de commande permettant le calcul des CI 50 et des CI 25 figurent dans le tableau O.2 ; cependant, on peut formater les équations permettant le calcul de toute CI p. Par exemple, les codes de commande de la CI 50 par le modèle logistique seraient les suivants :

```
nonlin
print          =     long
model drywts   =     t/(1+(0.25/0.75)*(logconc/x)^b)
save resid1/ resid
estimate/ start =    85, 0.6, 2 iter  =  200
use resid1
pplot residual
plot residual*logconc
plot residual*estimate
```

(3) Pour la colonne du tableau de données renfermant la variable à analyser, saisir l'en-tête dans la ligne intitulée : « model y= » (où *y* est la variable dépendante, par ex *jdrywt*).

(4) La 4e ligne du texte devrait se lire : « save resida/ resid », où *a* est le numéro auquel on affecte le fichier des résidus. Saisir ce même numéro à la 6e ligne (« use resida ») de sorte que l'on utilisera le même fichier pour produire un tracé de probabilité normale et des graphiques des résidus. Les lignes de commandes ci-après donnent des instructions pour produire un tracé de probabilité (« pplot residual »), un graphique des résidus en fonction de la valeur de la concentration ou du traitement (« plot residual*logconc ») et un graphique des résidus en fonction des valeurs prédites et ajustées (« plot residual*estimate »). Ces graphiques aident à évaluer les hypothèses de la normalité (par ex. tracé de probabilité) et d'homogénéité des résidus (par ex. graphiques des résidus) lorsqu'on évalue le modèle le mieux ajusté aux données (§ O.2.4).

(5) Dans la 5e ligne intitulée « estimate/start = », remplacer la moyenne des témoins et la CI p estimée (cf. tableau O.2 pour connaître les détails sur cette opération de remplacement relative à chaque modèle). Ces valeurs ont d'abord été obtenues par examen du nuage de points ou du graphique linéaire. Le modèle, dès qu'il converge, donnera un ensemble de paramètres à partir desquels on signale la CI p et ses limites de confiance à 95 % (c'est-à-dire le paramètre *x*). Il est essentiel de fournir des estimations exactes de chaque paramètre avant d'exécuter le modèle, sinon les itérations pourraient ne pas converger. L'estimation du paramètre d'échelle (tableau O.2) se situe habituellement entre 1 et 4. On peut modifier le nombre d'itérations, mais, dans l'exemple qui nous occupe, il a été réglé à 200 (c'est-à-dire « iter = 200 »). Typiquement, 200 itérations suffisent à un modèle pour le faire converger ; s'il en faut davantage, c'est probablement que l'on n'a pas utilisé le modèle qui convenait le mieux.

(6) Cliquer sur **File** (Fichier), puis **Submit Window** (Appliquer fenêtre) pour exécuter les commande ; on peut également cliquer avec le bouton droit de la souris et choisir **Submit Window** (Appliquer fenêtre). On produit

ainsi un imprimé des itérations, les paramètres estimés et une liste des données ponctuelles effectives avec les valeurs et résidus prévus correspondants. On se fait également présenter un graphique préliminaire de la droite estimée de régression. On devrait supprimer ce graphique, ce que l'on peut faire en sélectionnant le graphique se trouvant dans la fenêtre de gauche de l'écran principal. Sont également affichés des graphiques des résidus et un tracé de probabilité normale.

O.2.4 Examen des résidus et test d'hypothèses

L'examen des résidus de chaque modèle testé aide à déterminer si les hypothèses de la normalité de la distribution et de l'homoscédasticité sont vérifiées. Si l'on ne peut vérifier aucune des hypothèses, quel que soit le modèle examiné, on devrait consulter un statisticien pour obtenir des conseils sur l'emploi de modèles supplémentaires ou on devrait réanalyser les données par la méthode moins souhaitable qu'est l'interpolation linéaire (ICPIN ; § 6.4.2.2 ; annexe N).

O.2.4.1 Hypothèses de normalité

On devrait évaluer la normalité au moyen du *test de Shapiro-Wilk* décrit dans le § O.2.4.3 (v. aussi les § P.2.1 et P.2.2 de l'annexe P). Le tracé de probabilité normale, présenté dans l'**Output Pane** (Sous-fenêtre des résultats), peut également servir à déterminer si l'hypothèse de normalité est vérifiée. Les résidus devraient dessiner une ligne assez droite, traversant le graphique en diagonale ; la présence d'une courbe traduit un écart par rapport à la normalité. Le tracé de probabilité normale ne devrait cependant pas être le seul test de la normalité, parce que la décision concernant le degré de courbure dépendrait du jugement subjectif de l'utilisateur. Si les données n'obéissent pas à la loi normale, on devrait essayer un autre modèle, consulter un statisticien pour obtenir d'autres conseils ou analyser les données à l'aide de la méthode moins souhaitable qu'est l'interpolation linéaire.

O.2.4.2 Homogénéité des résidus

On devrait évaluer l'homoscédasticité (ou l'homogénéité) des résidus à l'aide du *test de Levene* d'après les consignes du § O.2.4.3 (v. aussi le § P.2.3 de l'annexe P) et par l'examen des graphiques des résidus. L'homogénéité des résidus se caractérise par une distribution égale de la variance des résidus, pour toutes les valeurs de la variable indépendante (fig. O.2A). Le test de Levene, s'il donne un résultat significatif, signifie que les données sont hétéroscédastiques, et l'on devrait alors examiner les graphiques des résidus. Si la variance varie de façon significative et que les graphiques des résidus ont nettement la forme d'un éventail ou d'un fuseau, on devrait répéter l'analyse des données au moyen de la régression pondérée. (Cf. la fig. O.2B montrant un tracé du « residual*estmate » ; un fuseau de direction opposée est également présenté dans le tracé du « residual*logconc ».) D'autre part, une divergence portant à croire à un manque systématique d'ajustement (fig. O.2C) signifie que l'on a retenu un modèle inadapté ou erroné.

O.2.4.3 Évaluation de la normalité et de l'homogénéité des résidus

La version 11.0 de SYSTAT peut appliquer les tests de Shapiro-Wilk et de Levene. On ne peut effectuer le test de Levene que si on applique une analyse de variance aux valeurs absolues des résidus calculés au § O.2.3.

(1) Cliquer sur **File** (Fichier), **Open** (Ouvrir), puis **Data** (Données), pour ouvrir le fichier de données renfermant les résidus créés au § O.2.3 (par ex. *resid1.syd*).

(2) Insérer un nouveau nom de variable dans une colonne vide, en cliquant deux fois sur le nom de la variable, ce qui fait apparaître la fenêtre '**Variable Properties**' (« Propriétés des variables »). Dans cette fenêtre, insérer un nom convenant aux résidus transformés (par ex. *absresiduals*) dans la zone intitulée '**Variable name:**' (« Nom de la variable »). Transformer les résidus en cliquant sur **Data** (Données), **Transform** (Transformer), puis **Let...** (Soit...). Ayant accédé à la fonction **Let...** (Soit...), choisir l'en-tête de colonne convenant aux données transformées (par ex. *absresiduals*), puis choisir **Variable** (Variable) dans la zone '**Add to**' (« Ajouter ») pour insérer la variable dans cette zone. Choisir la transformation appropriée (par ex. ABS pour la transformation des données en leur valeur absolue) dans la zone '**Functions:**' (« Fonctions ») [la zone '**Function Type:**' (« Type de fonction » devrait indiquer **Mathematical** [Mathématique]), puis choisir **Add** (Ajouter) pour insérer la

Tableau O.2. — Codes de commande dans SYSTAT pour les modèles de régression linéaire et non linéaire

Modèle	Codes de commande	Notes
exponentiel	nonlin print = long 'a') model mnlengths = a*exp(log((a-a*0.25-b*0.75)/a)*(logconc/x))+b save resid1/ resid estimate/ start = 25[a], 1[b], 0.3[c] iter = 200 use resid1 pplot residual plot residual*logconc plot residual*estimate	**Notes** [a] l'estimation de l'ordonnée à l'origine (c'est-à-dire *a*) [la réaction des organismes témoins] ; [b] le paramètre d'échelle (c'est-à-dire *b*) [valeur estimée entre 1 et 4) ; [c] l'estimation de la CI p pour l'ensemble de données (c'est-à-dire *x*).
Gompertz	nonlin print = long model mnlengths = g*exp((log(0.75))*(logconc/x)^b) save resid2/ resid estimate/ start = 16[a], 0.8[b], 1[c] iter = 200 use resid2 pplot residual plot residual*logconc plot residual*estimate	**notes :** [a] l'estimation de l'ordonnée à l'origine (c'est-à-dire *g*) [la réaction des organismes témoins] ; [b] l'estimation de la CI p pour l'ensemble de données (c'est-à-dire *x*) ; [c] le paramètre d'échelle (c'est-à-dire *b*) [valeur estimée entre 1 et 4).
hormétique	nonlin print = long model mnlengthr = (t*(1+h*logconc))/(1+((0.25+h*logconc)/ 0.75)*(logconc/x)^b) save resid3/ resid estimate/start = 48[a], 0.1[b], 0.7[c], 1[d] iter = 200 use resid3 pplot residual plot residual*logconc plot residual*estimate	**notes :** [a] l'estimation de l'ordonnée à l'origine (c'est-à-dire *t*) [la réaction des organismes témoins] ; [b] l'effet hormétique (c'est-à-dire *h*) [estimé entre 0,1 et 1] ; [c] l'estimation de la CI p pour l'ensemble de données (c'est-à-dire *x*). [b] le paramètre d'échelle (c'est-à-dire *b*) [valeur estimée entre 1 et 4) ;
linéaire	nonlin print = long model drywtr = ((-b*0.25)/x)*logconc+b save resid4/ resid estimate/start = 5[a], 0.7[b] iter = 200 use resid4 pplot residual plot residual*logconc plot residual*estimate	**notes :** [a] l'estimation de l'ordonnée à l'origine (c'est-à-dire *b*) [la réaction des organismes témoins] ; [b] l'estimation de la CI p pour l'ensemble de données (c'est-à-dire *x*).
logistique	nonlin print = long model drywts = t/(1+(0.25/0.75)*(logconc/x)^b) save resid5/resid estimate/start = 85[a], 0.6[b], 2[c] iter = 200 use resid5 pplot residual plot residual*logconc plot residual*estimate	**notes :** [a] l'estimation de l'ordonnée à l'origine (c'est-à-dire *t*) [la réaction des organismes témoins] ; [b] l'estimation de la CI p pour l'ensemble de données (c'est-à-dire *x*) ; [c] le paramètre d'échelle (c'est-à-dire *b*) [valeur estimée entre 1 et 4).

fonction dans la zone '**Expression:**' (« Expression ». Choisir l'en-tête de colonne renfermant les données d'origine non transformées (c'est-à-dire les résidus), puis **Expression** (Expression), dans la zone '**Add to**' (« Ajouter à »), pour insérer la variable dans la zone '**Expression:**' (« Expression »). Cliquer sur **OK**, ce qui fera apparaître les données transformées dans la colonne appropriée. Enregistrer les données.

(3) Pour effectuer le test de Shapiro-Wilk, cliquer sur **Analysis** (Analyse), **Descriptive Statistics** (Statistiques descriptives), puis **Basic Statistics...** (Statistiques de base...). La fenêtre '**Column Statistics**' (« Statistiques de colonne ») apparaît. Choisir les résidus de la zone '**Available variable(s):**' [« Variable(s) disponible(s) »], puis **Add** (Ajouter) pour insérer cette variable dans la zone '**Selected variable(s):**' [« Variable(s) sélectionnée(s) »]. Dans la zone '**Options**' (« Options »), choisir **Shapiro-Wilk normality test** (le test de normalité de Shapiro-Wilk), puis cliquer sur **OK**. Dans la fenêtre Outpout Organizer (organisateur des résultats) apparaîtra un petit tableau, où la valeur critique de Shapiro-Wilk (c'est-à-dire la 'SW Statistic' [statistique de Shapiro-Wilk]) et sa probabilité (c'est-à-dire la 'SW P-Value' [valeur *p* de S.-W.]) s'afficheront. Une valeur de probabilité supérieure au critère habituel de $p > 0{,}05$ dénote une distribution normale des données.

(4) Pour effectuer le test de Levene, cliquer sur **Analysis** (Analyse), **Analysis of Variance (ANOVA)** [Analyse de variance], puis **Estimate Model...** (Estimer le modèle...), ce qui fait apparaître la fenêtre '**Analysis of Variance: Estimate Model**' (« Analyse de variance : estimation du modèle ».

(5) Choisir la variable sous laquelle on veut grouper les données (par ex. *logconc*) et placer cette variable dans la zone '**Factor(s):**' [« Facteur(s) »] en cliquant sur **Add** (Ajouter).

(6) Choisir les résidus transformés (c'est-à-dire *absresiduals*), puis **Add** (Ajouter), pour insérer la variable dans la zone '**Dependent(s):**' [« Variable(s) dépendante(s) »]. Cliquer sur **OK**. Le résultat du test et un graphique des données apparaîtront dans la sous-fenêtre '**Output Pane**' (« Sous-fenêtre des résultats »). Une valeur de probabilité supérieure au critère habituel de $p > 0{,}05$ signifie que les données sont homogènes.

O.2.5 Pondération des données

Si, d'après le test de Levene, les résidus sont hétéroscédastiques et que, d'un traitement à l'autre, la variance varie de façon significative (c'est-à-dire disposition nettement en éventail ou en fuseau, fig. O.2B), il faudrait réanalyser les données par régression pondérée. Le facteur de pondération accordé à un traitement donné est l'inverse de la variance des observations correspondant à ce traitement. Dans la régression pondérée, on compare l'erreur type de la CI p (présentée dans SYSTAT comme l'erreur type asymptotique ['A.S.E.' (pour *asymptotic standard error*) ; v. fig. O.3]) à l'erreur calculée par régression non pondérée. Si les deux erreurs types diffèrent de plus de 10 %, on retient comme meilleur choix la régression pondérée. Cependant, si la variance correspondant à tous les traitements varie de façon significative et si les erreurs types des régressions pondérées et non pondérées[73] diffèrent de moins de 10 %, on devrait consulter un statisticien sur d'autres modèles ou on pourrait utiliser la méthode d'interpolation linéaire. Pour chacun des modèles retenus, on compare la régression pondérée et la non pondérée, tout en effectuant la sélection finale du modèle et de la méthode de régression. Par ailleurs, si le test de Levene révèle une non-homogénéité et que les graphiques des résidus montrent la non-divergence de ces derniers (par ex. fig. O.2C), on pourrait avoir retenu un modèle inadapté ou erroné. Ce serait encore l'occasion de consulter un statisticien sur des modèles de rechange.

73. La valeur de 10 % est purement empirique. Des tests permettent de juger objectivement de l'amélioration due à la pondération, mais ils dépassent notre propos. On ne devrait recourir à la pondération qu'en cas de nécessité, l'opération risquant de compliquer davantage la modélisation. On devrait consulter un statisticien lorsque la pondération est nécessaire, mais que les estimations résultantes des paramètres sont absurdes.

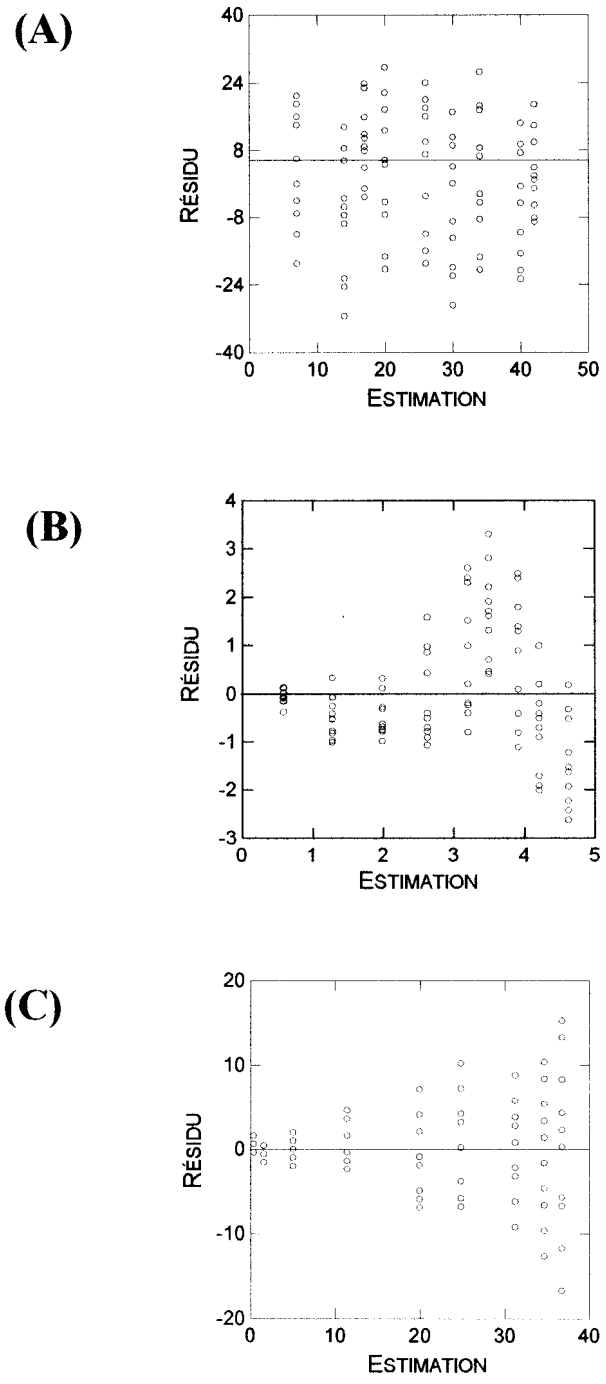


Figure O.2. — Résidus en fonction des valeurs prédites. Le graphique A dénote une homoscédasticité. Les graphiques B et C montrent deux types d'hétéroscédasticité : dans le premier cas, la répartition des points en éventail ou en fuseau nécessite un examen plus poussé à l'aide d'une régression pondérée ; dans le second cas, la répartition des points révèle un manque systématique d'ajustement, en raison du choix du mauvais modèle.

(1) Cliquer sur **File** (Fichier), **Open** (Ouvrir), puis **Data** (Données). Choisir le fichier renfermant l'ensemble de données à pondérer. Insérer les deux nouveaux noms de variables dans l'en-tête de colonne, en cliquant deux fois sur le nom d'une variable, ce qui ouvre la fenêtre '**Variable Properties**' (« Propriétés des variables »). Dans cette fenêtre, insérer le nom qui convient de la variable à laquelle on s'intéresse, choisir le type de variable et, si on le désire, ajouter des commentaires. Les deux nouveaux en-têtes de colonnes devraient indiquer la variance d'une variable particulière (par ex. *varjdrywt*) et l'inverse de la variance de cette variable (par ex. *varinvsjdrywt*). Enregistrer le fichier de données en cliquant sur **File** (Fichier), puis **Save** (Enregistrer).

(2) Cliquer sur **Data** (Données), puis sur **By Groups...** (Par groupe...). Cliquer sur la variable indépendante (c'est-à-dire *logconc*), puis sur **Add** (Ajouter), pour insérer la variable dans la zone '**Selected variable(s):**' [« Variable(s) choisie(s) »] ; cela permettra la détermination de la variance recherchée pour chaque traitement (c'est-à-dire par « groupe »). Cliquer sur **OK**.

(3) Cliquer sur **Analysis** (Analyse), **Descriptive Statistics** (Statistiques descriptives), puis **Basic Statistics...** (Statistiques de base...). Choisir la variable à pondérer (par ex. *jdrywt*), puis cliquer sur **Add** (Ajouter) pour l'insérer dans la zone '**Selected variable(s):**' [« Variable(s) choisie(s) »]. Cliquer sur **Variance** dans la zone '**Options**' [« Options »], puis sur **OK**. La variance recherchée, groupée par traitement, sera affichée dans la sous-fenêtre '**Output Pane**' (« Sous-fenêtre des résultats ») de l'écran principal.

(4) Cliquer sur **Data** (Données), **By Groups...** (Par groupes...), puis dans la boîte à côté de **Turn off** (Fermer), puis sur **OK** pour que toute analyse subséquente ne se fonde pas sur des traitements individuels, mais sur l'ensemble complet de données.

(5) Revenir au fichier de données en cliquant sur l'onglet '**Data Editor**' (« Éditeur de données ») de l'écran principal. Transférer les variances correspondant à chaque concentration ou traitement vis-à-vis la concentration correspondante de la colonne des variances (par ex. *varjdrywt*). À noter que la variance est la même entre les répétitions d'un même traitement.

(6) Cliquer sur **Data** (Données), **Transform** (Transformer), puis **Let...** (Soit...), puis, enfin, sur l'en-tête de colonne renfermant l'inverse de la variance (par ex. *varinvsjdrywt*) de la variable à laquelle on s'intéresse, puis sur **Variable** (Variable), dans la zone **'Add to**' (« Ajouter à ») pour insérer la variable dans la zone '**Variable:**' (« Variable »). Sélectionner la zone '**Expression:**' (« Expression »), puis l'en-tête de la colonne des variances (par ex. *varjdrywt*) de la variable à laquelle on s'intéresse, pour chaque répétition et concentration, puis cliquer sur **Expression** dans la zone '**Add to**' (« Ajouter à ») pour insérer la variable dans la zone '**Expression:**' (« Expression »). Cliquer sur **OK**. L'inverse de la variance de chaque répétition et concentration s'affichera dans la colonne appropriée. Enregistrer les données en cliquant sur **File** (Fichier), puis **Save** (Enregistrer).

(7) Cliquer sur **File** (Fichier), **Open** (Ouvrir), puis **Command** (Commandes) ; ouvrir le fichier des codes de commande pour l'estimation des paramètres de l'équation (par ex. § O.2.3, étape 2) du modèle retenu pour l'analyse *sans pondération.*

(8) Insérer une rangée supplémentaire après la 3^{e} ligne en tapant : « weight=varinvsy » où *y* est la variable dépendante à pondérer (par ex. weight=varinvsjdrywt), conformément à la 4^{e} ligne ci-dessous :

```
nonlin
print=long
model drywts = t/(1+(0.25/0.75)*(logconc/x)^b)
weight=varinvsdrywts
```

SYSTAT Rectangular file C:\SYSTAT\STATAPP.SYS, (Fichier rectangulaire SYSTAT C:\SYSTAT\STAPAPP.SYS)
created Tue May 25, 2004 at 13:46:14, contains variables: (créé le mardi 25 mai 2005 à 13:46:14, renferme les variables suivantes :)
CONC REP LOGCONC JUVENILES JDRYWT

Iteration (itération)

No.	Loss (perte)	G	X	B
0	.452080D+04	.340000D+02	.400000D+00	.100000D+01
1	.184579D+04	.328003D+02	.708478D+00	.157121D+01
2	.157417D+04	.331384D+02	.696189D+00	.197718D+01
3	.156445D+04	.329695D+02	.702780D+00	.211068D+01
4	.156432D+04	.329461D+02	.703292D+00	.212794D+01
5	.156432D+04	.329427D+02	.703387D+00	.212931D+01
6	.156432D+04	.329424D+02	.703394D+00	.212941D+01

Dependent variable is JUVENILES (variable dépendante : *juveniles* [jeunes])

Source (source)	Sum-of-Squares (somme des carrés)	df (diff.)	Mean-Square (moy. des carrés)
Regression (régression)	41208.68	3	13736.228
Residual (résidu)	1564.32	87	17.981
Total (total)	42773.00	90	
Mean corrected (moyenne corrigée)	15140.46	89	

→ moyenne des carrés des erreurs résiduelles

Raw R-square (1-Residual/Total) = 0.963 (R^2 brut [1 – résidu/total])
Mean corrected R-square (1-Residual/Corrected) = 0.897 (R^2 moy. corrigé [1 – résidu/corrigé])
R(observed vs predicted) square = 0.897 (R[observé vs prédit]^2)

Parameter (paramètre)	Estimate (valeur estimée)	A.S.E. (ETA)	Param/ASE (param./ETA)	Wald Confidence Interval (intervalle de confiance de Wald) Lower (inf.)	< 95 % >	Upper (sup.)
G	32.942	1.031	31.952	30.893		34.992
X	0.703	0.031	22.898	0.642		0.764
B	2.129	0.229	9.299	1.674		2.585

Case (cas)	JUVENILES (Jeunes) Observed (observé)	JUVENILES (Jeunes) Predicted (prédit)	Residual (résidu)
1	36.000	32.942	3.058
2	31.000	32.942	-1.942
3	22.000	32.942	-10.942
4	25.000	32.942	-7.942
5	39.000	32.942	6.058
6	42.000	32.942	9.058
[...]	[...]	[...]	[...]
86	2.000	0.337	1.663
87	0.000	0.337	-0.337
88	0.000	0.337	-0.337
89	1.000	0.337	0.663
90	0.000	0.337	-0.337

CI p, erreur type asymptotique (ETA) et limites inférieure et supérieure de confiance à 95 %

Asymptotic Correlation Matrix of Parameters (matrice de corrélation asymptotique des paramètres)

	G	X	B
G	1.000		
X	-0.696	1.000	
B	-0.611	0.566	1

Figure O.3. — Exemple des résultats initiaux donnés par le modèle de Gompertz dans la version 11 de SYSTAT. On y trouve la moyenne des carrés des erreurs résiduelles utilisée pour trouver le modèle à retenir, de même que les CI p, l'erreur type de l'estimation ainsi que les limites supérieure et inférieure de confiance à 95 %. Dans un souci de concision, nous avons délibérément écourté l'affichage du nombre de cas ; cependant les résultats présentés par SYSTAT exposent tous les cas ayant donné lieu à une mesure effective de la variable ainsi que l'estimation prédite et le résidu qui lui correspondent.

```
save resid2/ resid
estimate/ start = 85, 0.6, 2 iter=200
use resid2
pplot residual
plot residual*logconc
plot residual*estimate
```

(9) Attribuer un nouveau nombre aux résidus dans la ligne intitulée « save resida » (où *a* représente ce nombre).

(10) Insérer la moyenne des témoins et la CI p estimée dans la ligne intitulée « estimate/ start... » (cf. tableau O.2 pour connaître les détails de l'opération relatifs à chaque modèle). Ces estimations seront les mêmes que celles que l'on a utilisées pour l'analyse sans pondération.

(11) Cliquer sur **File** (Fichier), puis sur **Submit Window** (Appliquer fenêtre) pour exécuter les commandes. Cela produira le résultat des itérations, les paramètres estimés et une liste des données ponctuelles avec les résidus et les données ponctuelles prédites qui leur correspondent, tous dans la sous-fenêtre '**Output Pane**' (« Sous-fenêtre des résultats ») de l'écran principal. Un graphique préliminaire de la droite de régression estimée s'affichera également : il faudrait le supprimer. S'afficheront également un tracé de probabilité normale et des graphiques des résidus.

(12) Procéder à l'analyse décrite dans le § O.2.4 pour s'assurer de la confirmation de toutes les hypothèses du modèle.

(13) Comparer l'analyse de régression pondérée à la non pondérée. Choisir la pondérée si elle arrive à une erreur type de la CI p inférieure de 10 % à celle de la régression non pondérée.

O.2.6 Valeurs aberrantes et observations inhabituelles

Une valeur aberrante est une mesure qui ne semble pas s'accorder aux autres résultats d'un essai. On peut reconnaître les valeurs aberrantes et les observations inhabituelles à l'examen de l'ajustement de la courbe concentration-réponse à tous les points de données et à l'examen des graphiques des résidus. Si on découvre une valeur aberrante, on devrait suivre le conseil général donné dans le § 10.2, ce qui comprend l'examen de toutes les conditions expérimentales et des enregistrements relatifs à l'essai, électroniques ou manuels, pour y déceler une erreur humaine. Il faut que, pour tous les traitements, l'examen soit identique et non pas qu'il porte uniquement sur le traitement donnant lieu à l'anomalie. L'examen devrait aussi prendre en considération la variabilité biologique naturelle et d'autres causes biologiques de l'anomalie apparente. Si on reconnaît l'existence d'une anomalie, les analyses devraient être effectuées avec et sans la valeur aberrante. Peu importe l'analyse considérée comme définitive, il faut que le rapport final soit accompagné d'une description des données, des valeurs aberrantes et des deux analyses avec leurs conclusions interprétatives. S'il semble y avoir eu plus d'une observation aberrante, le modèle choisi devrait être réévalué quant à son à-propos et l'on devrait envisager des solutions de rechange.

La fonction ANOVA de SYSTAT peut être une façon de déterminer si les données renferment ou non des valeurs aberrantes. Cependant, cette fonction repose sur l'hypothèse selon laquelle les résidus obéissent à la loi normale, et il faut avoir vérifié cette hypothèse avant d'utiliser la fonction. La présence de valeurs aberrantes peut également être déterminée à partir des graphiques des résidus ainsi qu'au moyen de certains tests décrits dans le § 10.2.

(1) Effectuer une analyse de variance de la façon décrite dans le § O.4, afin de déterminer s'il se trouve des valeurs aberrantes parmi les données. Toute valeur aberrante sera identifiée par un numéro de cas correspondant au numéro de rang dans le fichier de données de SYSTAT. Ce programme utilise les résidus « studentisés » comme indicateurs de valeurs aberrantes ; des valeurs supérieures à 3 dénotent une possible aberrance. Cela devrait être confirmé par les graphiques des résidus.

(2) Si l'on veut effectuer une analyse sans la donnée anormale, supprimer cette dernière du tableau (fichier) de données originelles, puis enregistrer le fichier sous un *nouveau* nom (c'est-à-dire cliquer sur **File** [Fichier], puis **Save As...** [Enregistrer sous]). Par exemple, le nouveau nom du fichier pourrait contenir la lettre *o* (pour outlier[s] removed [valeurs aberrantes supprimées]) à la fin du nom original du fichier.

(3) Répéter la régression avec les données débarrassées des valeurs aberrantes, en utilisant le même modèle et les mêmes paramètres estimés que ceux qui ont été utilisés alors que les valeurs aberrantes étaient présentes. On pourrait également utiliser un modèle de rechange pour l'analyse s'il se traduisait par un ajustement meilleur et une plus petite moyenne des carrés des erreurs résiduelles. Si la suppression des valeurs aberrantes ne modifie pas sensiblement la moyenne des carrés des erreurs résiduelles et la CI p (y compris ses intervalles de confiance), l'analyste devrait utiliser son jugement professionnel pour déterminer quelle analyse est supérieure. Il doit motiver son choix d'analyse et produire les enregistrements des autres analyses.

O.2.7 Sélection du modèle le plus approprié

Une fois que tous les modèles parmi lesquels il faut choisir ont été ajustés, il faudrait évaluer chacun d'eux relativement à la normalité, à l'homogénéité des résidus et à la moyenne des carrés des erreurs résiduelles. On devrait retenir comme le plus approprié le modèle qui satisfait à toutes les hypothèses et auquel correspond la plus petite moyenne des carrés des erreurs résiduelles (cf. fig. O.3). Cependant, si plus d'un modèle aboutit à la même moyenne des carrés des erreurs résiduelles et où tous les autres facteurs sont équivalents, le meilleur choix serait le modèle le plus simple. La moyenne des carrés des erreurs résiduelles est présentée dans la sous-fenêtre '**Output Pane**' (« Sous-fenêtre des résultats ») immédiatement après les itérations et avant les estimations des paramètres. Si on a effectué des régressions pondérée et non pondérée, on devrait choisir la meilleure, conformément aux critères exposés dans le § O.2.5. Si aucun des modèles ne permet un ajustement convenable aux données, on devrait consulter un statisticien ou on devrait analyser les données par la méthode moins souhaitable de l'interpolation linéaire.

O.2.8 Tracé de la courbe concentration-réponse

Une fois le modèle approprié retenu, il faut tracer sa courbe concentration-réponse.

(1) Dans l'écran de l'éditeur de commandes au bas de l'écran, copier l'équation du modèle, prise parmi les codes de commandes servant au calcul des estimations pour le modèle retenu. C'est l'équation à droite du signe =, à la 3^{e} ligne des codes de commandes énumérés dans le tableau O.2. L'équation devrait comprendre les caractères alphabétiques originels (par ex. *t, b, h,* etc.). On peut copier l'équation en la mettant en surbrillance et en cliquant sur **Edit** (Édition), puis **Copy** (Copier) [ou en actionnant le bouton droit de la souris, puis en cliquant sur **Copy** (Copier)].

(2) Cliquer sur **File** (Fichier), **Open** (Ouvrir), puis **Command** (Commandes) et ouvrir un fichier existant de commande graphique (c'est-à-dire tout fichier ayant l'extension *.cmd) semblable à l'exemple ci-après (ou, au besoin, en créer un), à l'aide du modèle logistique. Le premier tracé (c'est-à-dire *plot*) est un nuage de points de la variable indépendante en fonction de la série de logarithmes de concentrations (log concentration). Le second tracé (c'est-à-dire *fplot*) est l'équation de régression, superposée au nuage de points.

```
graph
begin
plot drywts*logconc/ title = 'Dry Mass of Barley Shoots', xlab = 'Log(mg boric acid/kg soil d.wt)', ylab
= 'Mass (mg)',
xmax = 2, xmin = 0, ymax = 90, ymin = 0
fplot y = 80.741/(1+(0.25/0.75)*(logconc/0.611)^2.533); xmin = 0,
```

xmax = 2, xlab = '' ymin = 0, ylab = '', ymax = 90
end

(3) Coller l'équation copiée à la place de l'équation préexistante (figurant dans la zone grisée qui précède), en mettant l'équation précédente en surbrillance, puis en cliquant sur **Edit** (Édition), puis **Paste** (Coller) [ou en actionnant le bouton droit de la souris, puis en cliquant sur **Paste** (Coller)]. Remplacer tous les caractères alphabétiques (par ex. *t, b, h, x, a,* etc.) ainsi que les estimations respectives fournies dans la sous-fenêtre '**Output Pane**' (« Sous-fenêtre des résultats ») produite par l'application du modèle retenu.

(4) Saisir l'information convenable dans la ligne intitulée : « plot y*logconc... », où *y* est la variable dépendante à l'étude (par ex. *drywts*). Ajuster les valeurs numériques de « xmax » (c'est-à-dire la concentration logarithmique maximale utilisée) et « ymax » (cf. § O.2.1, étape 7) en conséquence. S'assurer que toutes les entrées de « xlab » et « ylab » (c'est-à-dire les étiquettes des axes) sont justes. Sinon les corriger en conséquence. S'assurer que tous les guillemets et toutes les virgules sont placés dans le programme de commande de la façon montrée dans l'exemple précédent ; SYSTAT est indifférent à la casse et à l'espacement.

title	s'applique au titre du graphique ;
xlab	s'applique au libellé de l'axe des abscisses (*x*) ;
xmin	s'applique à la valeur minimale demandée pour cet axe ;
xmax	s'applique à la valeur maximale demandée pour cet axe ;
ylab	s'applique au libellé de l'axe des ordonnées (*y*) ;
ymax	s'applique à la valeur maximale demandée pour cet axe ;
ymin	s'applique à la valeur minimale demandée pour cet axe.

Les valeurs de *xmin*, *xmax*, *ymin* et *ymax* doivent être les mêmes dans les deux tracés, pour que la superposition de la droite de régression sur le nuage de points soit parfaite. Un exemple du graphique final de régression est reproduit dans la fig. O.1, pour chacun des cinq modèles proposés.

(5) Cliquer sur **File** (Fichier), puis **Save As** (Enregistrer sous) afin d'enregistrer les codes de commande graphique dans le dossier approprié de travail utilisant le même codage que celui qui a servi à produire le fichier de données, avec indication du modèle auquel correspond la régression. Cliquer sur **Save** (Enregistrer) pour enregistrer le fichier.

(6) Cliquer sur **File** (Fichier), puis **Submit Window** (Appliquer fenêtre) pour traiter les codes de commande. Apparaîtra un graphique de la régression utilisant les paramètres estimés pour le modèle retenu.

O.3 Détermination de CI p supplémentaires

Dans certains cas, il pourrait être souhaitable d'estimer une deuxième CI p avec une autre valeur de *p*. Bien que le paragraphe qui suit et la fig. O.1 concernent la détermination de la CI 25, on peut adapter les modèles à toute valeur de *p* (par ex. la CI 20).

(1) Cliquer sur **File** (Fichier), **Open** (Ouvrir), puis **Command** (Commandes) et ouvrir le fichier correspondant aux codes de commande employés pour produire les estimations des paramètres (cf. tableau O.2 pour un aperçu des codes de commande de chaque modèle). Modifier l'équation du modèle en vue du calcul de la CI p recherchée (par ex. la CI 25). La fig. O.1 renferme des conseils sur la modification des modèles pour permettre le calcul de la CI 25. On peut déterminer toute CI p en modifiant les fractions utilisées dans chaque modèle. Par exemple, pour calculer la CI 20 au moyen du modèle logistique, il faudrait remplacer l'équation servant au calcul de la CI 50 ('t/[1 + (logconc/x)^b]) par la suivante : t/[1(0,20/0,80)*(logconc/x)^b.

(2) Une fois l'équation ajustée pour la CI p à laquelle on s'intéresse, suivre chaque étape exposée dans le § O.2.3. Toutefois, remplacer l'estimation initiale de la CI p dans 5e ligne, intitulée « estimate/ start= » (cf. fig. O.1 pour connaître les détails du remplacement dans chaque modèle). C'est la valeur découlant, à l'origine, d'un examen du nuage de points ou d'un graphique linéaire. Le modèle, dès qu'il converge, donne un ensemble de paramètres parmi lesquels sont signalés la CI p et ses limites de confiance au seuil de 95 % (c'est-à-dire le paramètre *x*).

(3) Passer à l'analyse décrite dans les § O.2.4 à O.2.8.

O.4 Analyse de variance (ANOVA)

(1) Cliquer sur **File** (Fichier), **Open** (Ouvrir), puis **Data** (Données) pour ouvrir le fichier de données renfermant toutes les observations concernant l'ensemble de données à l'examen.

(2) Cliquer sur **Analysis** (Analyse), **Analysis of Variance (ANOVA)** [Analyse de variance], puis **Estimate Model...** (Estimer le modèle...).

(3) Sélectionner la variable sous laquelle il faut grouper les données (par ex. logconc) et placer cette variable dans la zone '**Factor(s):**' [« Facteur(s) »] en cliquent sur **Add** (Ajouter).

(4) Choisir la variable à laquelle on s'intéresse (par ex. jdrywt), puis cliquer sur **Add** (Ajouter), afin d'insérer la variable dans la zone '**Dependent(s):**' [« Variable(s) dépendante(s) »].

(5) Sélectionner la zone à côté de '**Save**' (« Enregistrer ») [à l'angle inférieur gauche de la fenêtre 'Analysis of Variance: Estimate Model' (« Analyse de variance : estimer le modèle ») puis la faire défiler vers le bas jusqu'aux sélections d'accompagnement pour choisir **Residuals/Data** (Résidus/données). Saisir un nom convenable de fichier dans la zone vide adjacente pour sauvegarder (enregistrer) les résidus (par ex. anova1). Cliquer sur **OK**. Apparaît un graphique des données et les résultats produits, dans la sous-fenêtre '**Output Pane**' (« Sous-fenêtre des résultats »). Toute valeur aberrante, d'après les résidus « studentisés », est alors identifiée (v. § O.2.6 pour ce qui concerne les valeurs aberrantes).

(6) Évaluer les hypothèses de normalité et d'homogénéité des résidus, conformément au § O.2.4, à l'aide du fichier de données créé pour enregistrer les résidus ou les données avant la réalisation de l'analyse de variance (c'est-à-dire anova1). Effectuer les évaluations à l'aide des tests de Shapiro-Wilk et de Levene. On peut utiliser le codage ci-dessous pour examiner les graphiques des résidus :

```
graph
use anova1
plot residual*logconc
plot residual*estimate.
```

Test d'hypothèse(s)

P.1 Méthodes statistiques

Par le passé, on a fréquemment utilisé le test d'hypothèse(s) à l'égard des effets quantitatifs sublétaux tels que la taille atteinte. Il est possible de transformer des données quantiques en données quantitatives, analysables au moyen du test d'hypothèse(s) [§ 2.92 et 2.9.3]. On peut appliquer directement le test d'hypothèse(s) à des données quantiques, sans difficultés statistiques, si le nombre d'observations dans une répétition est d'au moins 100, parce que les données deviennent semblables à des distributions quantitatives. Par exemple, dans l'essai avec des œufs d'oursins, on compte les œufs fécondés, parmi les 100 ou 200 premiers qui se trouvent sur une lame de verre. La méthode d'essai d'Environnement Canada (1992f) reconnaît la nature quantique de l'effet, mais les nombres en cause sont suffisamment grands pour qu'on l'assimile à un effet quantitatif. Cette marche à suivre n'est pas recommandée pour de petits nombres d'observations dans chaque répétition, 40 par ex. L'importance des grands nombres réside dans le fait que le saut quantique de l'effet causé par un individu réagissant à l'intérieur d'un groupe de 100 ne représente que 1 %, ce qui s'approche d'une distribution continue et est satisfaisant pour les techniques quantitatives.

Dans TOXSTAT (1996 ; WEST et Gulley, 1996) et CETIS (2001), on présente des méthodes statistiques pour le test d'hypothèse(s) et on les explique avec des conseils à l'appui dans USEPA (1994a), Newman (1995) ainsi que dans diverses méthodes d'essai de toxicité sublétale d'Environnement Canada. Les logiciels TOXSTAT et CETIS sont vendus dans le commerce, et d'autres fournisseurs proposent des programmes généraux élargis d'analyse informatisée. Il faudrait suivre les consignes figurant dans la notice du programme. Tous les fournisseurs de progiciels modifient plus ou moins les procédures dans les versions successives du logiciel.

Une échelle logarithmique est importante pour le choix des concentrations expérimentales ; cependant, il est inutile de s'assurer que l'on utilise des logarithmes de la concentration dans l'estimation de la CSEO et de la CEMO. Les logarithmes n'entrent pas dans l'analyse statistique, parce que les comparaisons statistiques se font entre les *effets* observés. On pourrait tout aussi bien identifier les groupes en utilisant des nombres arbitraires, des lettres ou des appellations. Dans certains cas, on tient compte de la concentration, par ex. le test de Williams tient compte de l'ordre des concentrations, mais non, cependant, de leur grandeur absolue.

P.2 Tests de la normalité et de l'homogénéité de la variance

P.2.1 Test de normalité de Shapiro-Wilk

Pour ce test, les calculs sont compliqués et ils seraient fastidieux si on les faisait à la main. Le programme TOXSTAT et d'autres programmes informatiques les effectuent rapidement. Les étapes mathématiques sont exposées dans Newman (1995) et dans un exemple présenté dans USEPA (1995). La dernière étape est la comparaison avec une valeur critique (W) trouvée dans des tables (Shapiro et Wilk, 1965 ; D'Agostino, 1986). Pour ce test, la taille minimale de l'échantillon est de 3.

On peut donner un exemple de test de la normalité d'après les données du tableau P.1. Les données représentent les gains de poids dans des groupes d'alevins de truite arc-en-ciel se trouvant à la fin du stade vésiculé, exposés à diverses concentrations de cuivre jusqu'au début du stade de la truitelle nageant librement. On a employé cinq concentrations et un témoin. À chaque concentration, il y avait 12 poissons, bien que 3 fussent morts à la concentration maximale. Ces données réelles sont du laboratoire de Beak International, Inc. de Brampton (Ont.).

Dans le tableau P.1, les deux colonnes intitulées « Gain de poids » et « Résidu » sont utiles au test de Shapiro-Wilk. Chaque valeur d'un résidu est simplement le poids moyen du groupe, soustrait du poids individuel (v. le glossaire), et ces résidus sont les valeurs qui sont analysées par le test.

Tableau P.1. — Tableau de présentation des données sur la toxicité utilisé comme exemple de l'évaluation de la normalité. Les données représentent le gain de poids d'alevins vésiculés de truite arc-en-ciel exposés à du cuivre dans une eau dont la dureté est de 135 mg/L. Dans cet exemple, il n'y a pas de répétitions, mais, dans le test d'hypothèses, il y aurait toujours des répétitions. Données fournies par Beak International, Inc.

Cuivre (µg/L)	Gain de poids (mg)	Résidu (mg)	Rang dans le groupe	Proportion cumulative	Probit
Témoin	66,7	-43,9	1	0,0769	3,5738
	101,5	-9,1	2	0,1538	3,9797
	102,7	-7,9	3	0,2308	4,2638
	103,7	-6,9	4	0,3077	4,4976
	105,0	-5,6	5	0,3846	4,7066
	109,3	-1,3	6	0,4615	4,9034
	111,7	1,1	7	0,5385	5,0967
	112,6	2,0	8	0,6154	5,2930
	122,2	11,6	9	0,6923	5,5018
	125,7	15,1	10	0,7692	5,7362
	128,9	18,3	11	0,8462	6,0203
	137,3	26,7	12	0,9231	6,4262
moyenne	110,6				
12	64,0	-25,4	1	0,0769	3,5738
	67,3	-22,1	2	0,1538	3,9797
	81,8	-7,6	3	0,2308	4,2638
	85,6	-3,8	4	0,3077	4,4976
	85,8	-3,6	5	0,3846	4,7066
	92,0	2,6	6	0,4615	4,9034
	92,0	2,6	7	0,5385	5,0967
	92,1	2,7	8	0,6154	5,2930
	96,5	7,1	9	0,6923	5,5018
	96,6	7,2	10	0,7692	5,7362
	105,4	16,0	11	0,8462	6,0203
	114,1	24,7	12	0,9231	6,4262
moyenne	89,4				
25	51,5	-41,3	1	0,0769	3,5738
	73,4	-19,4	2	0,1538	3,9797
	80,2	-12,6	3	0,2308	4,2638
	81,5	-11,3	4	0,3077	4,4976
	88,3	-4,5	5	0,3846	4,7066
	88,6	-4,2	6	0,4615	4,9034
	91,7	-1,1	7	0,5385	5,0967
	96,4	3,6	8	0,6154	5,2930
	109,0	16,2	9	0,6923	5,5018
	109,1	16,3	10	0,7692	5,7362
	112,6	19,8	11	0,8462	6,0203
	131,5	38,7	12	0,9231	6,4262
moyenne	92,8				
48	54,6	-28,1	1	0,0769	3,5738
	56,4	-26,3	2	0,1538	3,9797
	57,7	-25,0	3	0,2308	4,2638
	78,0	-4,7	4	0,3077	4,4976
	79,6	-3,1	5	0,3846	4,7066
	80,8	-1,9	6	0,4615	4,9034
	81,9	-0,8	7	0,5385	5,0967
	83,3	0,6	8	0,6154	5,2930
	97,4	14,8	9	0,6923	5,5018
	106,4	23,8	10	0,7692	5,7362
	107,8	25,1	11	0,8462	6,0203
	107,9	25,3	12	0,9231	6,4262
	82,7				
65	49,8	-16,1	1	0,0769	3,5738
	54,9	-10,9	2	0,1538	3,9797
	56,1	-9,7	3	0,2308	4,2638
	56,4	-9,4	4	0,3077	4,4976
	60,2	-3,6	5	0,3846	4,7066
	62,6	-3,2	6	0,4615	4,9034
	64,0	-1,8	7	0,5385	5,0967
	65,0	-0,8	8	0,6154	5,2930
	68,8	3,0	9	0,6923	5,5018
	69,2	3,4	10	0,7692	5,7362
	81,2	15,4	11	0,8462	6,0203
	102,0	36,2	12	0,9231	6,4262
	65,9				
91	11,7	-25,4	1	0,0769	3,5738
	13,5	-22,1	2	0,1538	3,9797
	19,1	-20,0	3	0,2308	4,2638
	41,3	-2,2	4	0,3077	4,4976
	45,6	6,5	5	0,3846	4,7066
	47,4	8,3	6	0,4615	4,9034
	56,6	17,5	7	0,5385	5,0967
	57,2	18,1	8	0,6154	5,2930
	59,2	20,1	9	0,6923	5,5018
	39,1				

Les calculs aboutissent à une valeur critique *W* de 0,9836, et la valeur de la probabilité associée est 0,5, ce qui est très haut. Comparativement au critère habituel de $p > 0,05$, il est clair que les données suivent la loi normale. Pour une appréciation visuelle de ces données, v. la fig. P.1.

L'expérimentateur peut évaluer le degré de non-conformité par la *p*-valeur offerte dans le programme informatique ou, si nécessaire, dans une table des valeurs critiques de *W,* qui devrait donner les divers niveaux de probabilité à partir de 0,01 en montant. On peut s'attendre à des valeurs d'environ 0,3 à 1,0 comme résultats (*W*) du test de Shapiro-Wilk, la valeur inférieure signifiant qu'il existe un écart considérable par rapport à la normalité, tandis que la valeur de 1,0 signifie presque aucun écart.

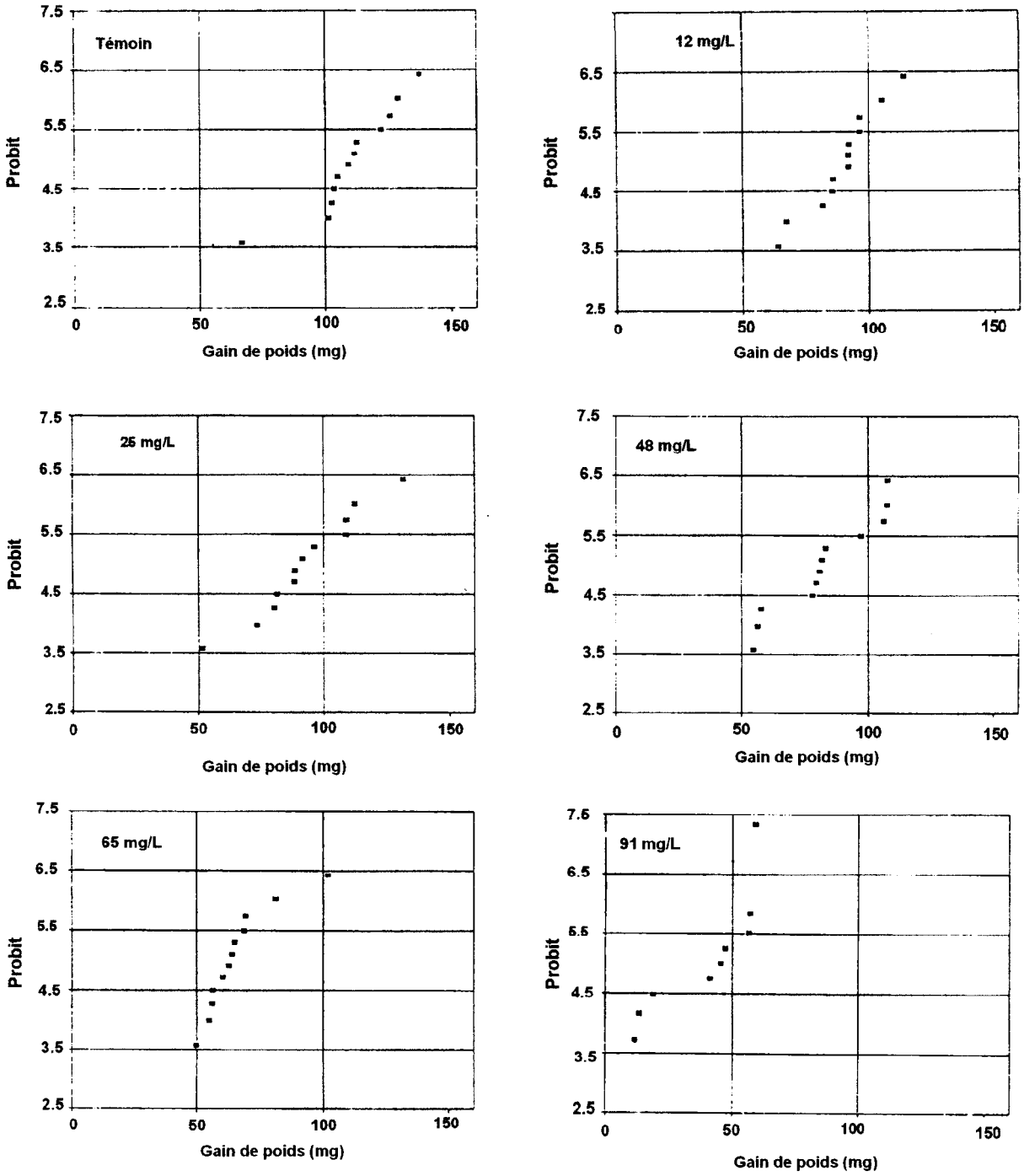


Figure P.1. — Graphiques permettant d'examiner la normalité apparente de la distribution des gains de poids d'alevins vésiculés de truite arc-en-ciel exposés à diverses concentrations de cuivre. Chaque graphique représente le rang cumulatif du gain de poids de chaque alevin dans la distribution de 12 alevins (sur une échelle verticale de probabilité), en fonction des gains de poids absolus (sur une échelle arithmétique). Trois alevins exposés à la concentration maximale sont morts.

Bien que les tests de normalité puissent porter sur les poids correspondant à chaque traitement, cela n'est pas recommandé. Les petites tailles de l'échantillon réduisent la puissance du test et augmentent la probabilité d'une erreur de première espèce.

P.2.2 Tracé d'un graphique pour vérifier la normalité

On recommande le test de Shapiro-Wilk (§ P.2.1) pour évaluer la normalité, et ce test devrait être le critère d'acceptation des données. En outre, il pourrait être instructif de tracer des graphiques de la distribution des données pour apprécier visuellement cette distribution. Les graphiques devraient se fonder sur les données originelles d'une répétition ou d'une concentration. Dans les cas où les données n'obéissent pas à la loi normale ou ne sont pas homogènes, le graphique pourrait en révéler la cause apparente. Il n'est pas recommandé de se fonder uniquement sur l'analyse des graphiques pour juger de la normalité des résultats parce qu'il faut, à cette fin, des techniques graphiques particulières, de même que de l'expérience et des compétences pour l'interprétation subjective. Dans les échantillons de petite taille, il pourrait survenir des changements brusques, qui pourraient facilement mener à la surinterprétation. Si on effectue une évaluation visuelle, les méthodes privilégiées sont, dans l'ordre, les diagrammes des quantiles, les boîtes à moustaches ou les diagrammes tiges et feuilles et les histogrammes.

En dépit de ces mises au point, on trouve, dans les publications, un appui en faveur de l'évaluation graphique de la normalité. L'appui de l'OCDE (OECD, 2004) est décrit dans le § 7.3.2 (note 54). Newman (1995) décrit brièvement la méthode et renvoie à des exemples détaillés dans Sokal et Rohlf (1981) et Miller (1986). Newman (1995) cite Miller qui aurait écrit : « Si un écart par rapport à la normalité ne peut pas se voir à l'œil sur papier probit, il ne vaut pas la peine de s'en soucier. » On ne contestera pas cependant que l'œil qui décèle cette anomalie doit être expérimenté.

On peut donner des exemples de graphiques ainsi produits avec les données du tableau P.1 (fig. P.1). Nous savons déjà que les données obéissent à la loi normale avec une forte valeur de probabilité, d'après les résultats du test de Shapiro-Wilk du § P.2.1 ; ainsi les graphiques de la fig. P.1 représentent des données relativement bonnes. Il faut souligner que le *test de normalité porte sur la distribution normale des résidus.* Bien que, en théorie, si les effets suivent la loi normale à chaque concentration, les résidus devraient également avoir une distribution normale, les tests de normalité devraient effectivement porter sur les résidus. En conséquence, la fig. P.1 ne représente pas l'évaluation visuelle à laquelle nous faisions allusion deux alinéas plus haut (diagrammes des quantiles, etc.) ; la fig. P.1 représente simplement ce à quoi des données relativement bonnes ressemblent sur des diagrammes probit.

Ci-dessous, on trouvera le plan des calculs et du traçage des graphiques. On utilise les trois dernières colonnes du tableau P.1. Dans d'autres sortes de tests, on pourrait remplacer le « gain de poids » par n'importe quel type de mesure.

- Pour chaque concentration (ou chaque répétition, le cas échéant), énumérer les mesures dans l'ordre croissant (Dans ce cas, les mesures seraient le gain de poids de chacun des jeunes poissons).

- Attribuer un numéro de rang, sur 12, à chaque gain de poids. Pour les valeurs égales, utiliser la moyenne des rangs.

- Pour chaque gain de poids, calculer la proportion cumulative des données représentées. Calculer ces valeurs en posant l'existence d'une valeur supplémentaire (12 + 1 = 13 pour la plupart des traitements du tableau P.1 et 9 + 1 pour la concentration maximale).

  La proportion cumulative = (rang du gain de poids) / (nombre de mesures + 1,0).

- Porter chaque proportion cumulative sur une échelle probit en fonction de son gain de poids. (Sinon, pour chaque proportion cumulative, obtenir le probit à l'aide d'un programme informatique ou le tirer d'une table et porter le probit sur une échelle arithmétique comme dans le tableau P.1 et la fig. P.1.)

Dans la fig. P.1, le gain de poids des truitelles présente une relation passablement linéaire dans la plupart des cas, ce qui porte à croire en l'existence d'une distribution probablement normale. On constate de petits écarts modérés par rapport à la normalité, particulièrement chez les individus présentant le moins de gain de poids dans le groupe témoin ainsi que chez les individus montrant le plus grand gain de poids (à 65 et 91 µg/L). Néanmoins, ces données se sont révélées avoir un haut degré de probabilité d'après le test de Shapiro-Wilk, de sorte que la fig. P.1 représente une normalité acceptable de la distribution.

Si les expériences dont rend compte le tableau P.1 étaient destinées à un test d'hypothèse(s), il y aurait des répétitions. Il y aurait un groupe supplémentaire de 12 alevins vésiculés dans une enceinte expérimentale séparée pour chaque répétition d'une concentration. Pour tracer le graphique ou tester la normalité des résidus, il faudrait représenter séparément chaque répétition sur le graphique.

Dans certains cas, une observation répétée serait un nombre unique tel que le poids total ou le poids moyen de tous les individus d'une enceinte expérimentale, ce qui est le cas du poids des larves dans l'essai employant des têtes-de-boule. Pour les essais ayant ce plan d'expérience, le poids moyen d'une répétition donnée serait ordonné parmi tous les poids moyens calculés pour la même concentration. Les résidus de ces classements et de ces poids moyens seraient portés sur un graphique. S'il n'y avait que deux ou trois valeurs, le graphique ne révélerait pas grand chose, et, de fait, l'exercice pourrait induire en erreur. Le test de Shapiro-Wilk resterait le critère.

P.2.3 Tests de l'homogénéité de la variance

La méthode que nous recommandons pour évaluer les équivariances est le test de Levene (1960), décrit dans Snedecor et Cochran (1980), mais qui, actuellement, ne fait pas partie des progiciels conçus pour l'écotoxicologie. Le test de Bartlett (1937) est usuel dans les progiciels, mais il présente un inconvénient (voir le texte qui suit). Le test d'O'Brien (1979) est quelque peu supérieur au test de Levene, mais est également absent des progiciels actuels de statistique. Les données fondées sur des proportions ne devraient pas être assujetties à ces tests.

Tous ces tests permettent de déterminer si les variances sont égales pour tous les traitements, l'hypothèse nulle étant l'absence de différence. Si les variances diffèrent notablement d'un traitement à l'autre, l'hypothèse de l'homogénéité dont on a besoin pour une analyse subséquente de variance est invalide. Les tests de variance partent de l'hypothèse selon laquelle les observations obéissent à la loi normale.

Le test de Bartlett est offert dans la plupart des logiciels d'écotoxicologie et il est largement utilisé. La statistique du test est calculée à partir des variances « intra-traitement » et des variances résiduelles. La comparaison finale concerne une valeur critique de khi-deux, pour le nombre approprié de degrés de liberté et une valeur choisie de probabilité **(α)**. Pour les échantillons dont la taille est inférieure à 5 individus, on utilise une table spéciale des valeurs critiques. La plupart des expérimentateurs laisseront le soin des calculs au programme informatique. Les étapes effectivement suivies sont montrées dans des exemples donnés par Newman (1995) et l'USEPA (1995).

Le test de Bartlett est excessivement sensible, si les données ne suivent pas la loi normale et, particulièrement, si les distributions sont asymétriques. L'ensemble de données pourrait alors être rejeté, à tort, par le test d'homogénéité de la variance.

Le test de Levene permet d'éviter ce problème, grâce à l'emploi de la moyenne des écarts absolus d'une observation par rapport à la moyenne du traitement auquel elle appartient, plutôt que la moyenne des écarts élevés au carré des variances « intra-traitement » et résiduelles. Comme nous l'avons mentionné, le test de Levene n'est pas un test usuel des progiciels et il n'est pas mentionné ni décrit dans certains manuels (Zar, 1999 ; Newman, 1995). La méthode de Levene pourrait cependant être mise en œuvre par un traitement manuel des données. On pourrait enregistrer chaque observation comme étant l'écart absolu par rapport à la moyenne « intra-traitement ». On effectuerait ensuite une

analyse de variance sur les observations enregistrées. Le test F pour la différence dans les observations enregistrées serait un test de l'hypothèse de l'homogénéité.

Le test d'O'Brien est quelque peu supérieur à celui de Levene par certains aspects techniques relevant des mathématiques. Cependant, il est même moins facilement accessible que celui de Levene et il n'est pas expliqué dans les manuels usuels (Snedecor et Cochran, 1980 ; Zar, 1999 ; Newman, 1995).

Si les données que l'on soumet au test sont des proportions, les variances différeront selon la proportion et, en conséquence, selon le traitement. On devrait analyser de telles données quantiques par des méthodes plus appropriées que le test d'hypothèse(s) [section 4] sinon on devrait les transformer de façon convenable (§ 2.9.3). L'USEPA (1994d) a lancé un avertissement sur une difficulté particulière que pose le test de l'homogénéité de la variance appliqué à des données proportionnelles [74], mais l'avertissement est sans objet si le test d'hypothèse(s) ne s'applique pas à des effets proportionnels.

P.2.4 Robustesse de l'analyse paramétrique et décisions sur son emploi

Si les données réussissaient le test de Shapiro-Wilk's et celui de Levene ou celui de Bartlett, l'analyse devrait se poursuivre avec des méthodes paramétriques, c'est-à-dire l'analyse de variance.

Si les données présentent des incohérences et ne satisfont pas à l'un ou à l'autre de ces tests, on pourrait les transformer statistiquement pour qu'elles satisfassent aux exigences de l'analyse. Il faudrait éviter la transformation, si c'est possible, parce que l'opération entraîne des complications et des inconvénients, décrits dans le § 2.9.2. Si on se décide pour la transformation, on soumet de nouveau l'ensemble des données modifiées aux tests de normalité et d'homogénéité, pour voir si, désormais, elles satisfont aux exigences. Dans l'affirmative, l'analyse pourrait se poursuivre par les méthodes paramétriques usuelles.

Si, même après transformation, les données ne peuvent pas satisfaire à aucun de ces tests concernant la distribution des données, alors l'analyse doit se faire par des méthodes non paramétriques (fig. 19). Les progiciels posent habituellement que l'analyse non paramétrique sera la seule option, lorsque l'un des tests de qualification a échoué.

Cependant, on peut faire valoir que l'analyse de variance et les tests subséquents de comparaisons multiples sont plutôt robustes, dans l'éventualité de petits écarts à la normalité et à l'homogénéité. Les tests ayant ces caractéristiques fonctionnent bien avec de grands échantillons, mais ils pourraient ne pas bien se comporter avec les petits échantillons que l'on trouve souvent dans les essais sur l'environnement. Le test de normalité peut être trop sensible si les variances ne sont pas égales, et vice versa [75].

---

74. Si l'expérimentateur avait décidé d'analyser directement des données proportionnelles (quantiques) par un test d'hypothèse(s), il devrait régler une situation. Celle-ci entraînerait un rejet inutile des résultats d'un test paramétrique par suite du test d'homogénéité de la variance par les test de Bartlett ou de Levene. On peut invoquer à cet égard l'exemple l'analyse de la fécondation des œufs d'oursins. Il se pourrait que dans chaque répétition du groupe témoin, la fécondation soit de 100 %. De même, il se pourrait que le taux de fécondation dans chaque répétition de la concentration maximale soit nul. Dans un cas comme dans l'autre, la variance de ce traitement serait nulle également. Dans le test d'homogénéité, la variance nulle entraînerait le rejet de l'hypothèse des équivariances. Dans cette éventualité, le traitement correspondant à la variance nulle devrait être *omis* du test de Bartlett ou de Levene, et on devrait adopter l'estimation consécutive de la variance « intra-traitement » (USEPA, 1994d). Si les autres traitements satisfaisaient à la condition de l'équivariance, on pourrait passer à l'analyse paramétrique. Dans les analyses subséquentes (analyse de variance et test de comparaisons multiples), on devrait utiliser tous les traitements, y compris ceux qui correspondraient aux effets nuls et de 100 %.

75. Le test de normalité de Shapiro-Wilk est sensible aux variances inégales, tandis que le test de Bartlett, habituellement recommandé en cas de variances inégales, est réputé sensible à la non-normalité. Vu cette sensibilité réciproque, l'expérimentateur pourrait être quelque peu justifié de ne pas considérer comme un dogme cette suite de décisions préalables aux tests.

La robustesse relative de l'analyse de variance a été décrite par Zar (1974)[76]. Newman (1995) a cité les travaux selon lesquels l'analyse de variance produit des probabilités réalistes si la distribution des données est au moins symétrique et si les variances des traitements sont moins du triple les unes des autres. Un programme statistique énonce que : « L'analyse de variance peut être valide même si on s'écarte de la normalité, particulièrement quand le nombre de répétions par groupe est élevé. Si les répétitions sont égales ou presque égales, l'hétérogénéité de la variance influe peu sur l'analyse. » (TOXSTAT, 1996). Des documents récents, publiés par l'USEPA semblent aussi montrer un adoucissement sur cette question, car on y lit, par ex., que : « Si les tests échouent..., une méthode non paramétrique... peut être plus appropriée. Cependant la décision... peut relever du jugement, et l'on devrait consulter un statisticien pour le choix de la méthode d'analyse. » (USEPA, 1995).

En conséquence, si les tests statistiques de normalité et d'homogénéité de la variance révèlent un écart léger à modéré par rapport aux exigences (c'est-à-dire échec marginal d'un test), l'expérimentateur pourrait vouloir consulter un statisticien sur l'éventuelle utilité de tests paramétriques.

Dans cette situation, certaines méthodes d'essai de toxicité sublétale d'Environnement Canada recommandent à la fois une analyse paramétrique et non paramétrique, la plus sensible des deux (concentration plus faible) donnant les estimations finales de la toxicité[77]. Nous recommandons cette marche à suivre, et les constatations obtenues par les deux méthodes devraient être signalées. On devrait présenter les résultats du test de Shapiro-Wilk et d'O'Brien (ou de Bartlett) ainsi qu'un graphique des résultats bruts.

P.3 Analyse de variance

Comme test paramétrique, on effectue une *analyse de variance*, dont l'objectif est double : d'abord voir s'il existe une différence globale entre toutes les valeurs moyennes prises deux à deux (ou plus) pour les divers traitements (concentrations). À cette fin, on teste l'*hypothèse nulle* (H_0) selon laquelle il n'existe aucune différence significative entre les valeurs moyennes des traitements. Si on trouve une différence, le second objectif de l'analyse de variance est d'obtenir une estimation de la *variance de l'erreur* ; celle-ci servira dans des tests ultérieurs visant à trouver les concentrations particulières qui diffèrent.

L'analyse de variance se sert de : *a*) la variance totale de l'essai ; *b*) la variance entre les concentrations ; *c*) la variance « intra-concentration » (c'est-à-dire entre les répétitions). Les estimations de la variance sont la « moyenne de la somme des carrés des écarts » (l'expression complète est *moyenne arithmétique des carrés des écarts à la moyenne*), d'habitude appelées *erreur quadratique moyenne*. On les obtient en divisant la somme des carrés des écarts à la moyenne par le nombre de degrés de liberté. La somme des carrés des écarts à la moyenne s'obtient par soustraction de chaque observation (répétition) de la moyenne de la catégorie (concentration), élévation de cette différence au carré et sommation de tous les carrés. Le nombre de degrés de liberté est le nombre d'éléments dans la catégorie moins 1.

Les valeurs pertinentes produites par l'analyse sont présentées dans le tableau P.2. Ces valeurs hypothétiques correspondraient à un essai employant 5 concentrations, à raison de trois enceintes (répétitions) par concentration[78].

---

76. « L'expérience a montré que les analyses de variance et les test *t* sont habituellement assez robustes pour bien fonctionner, même si les données s'écartent quelque peu des conditions de la normalité, de l'homoscédasticité et de l'additivité. Mais des écarts graves peuvent mener à des conclusions fausses. » (Zar, 1974).

77. La justification de cela se fonde apparemment sur l'hypothèse selon laquelle de nombreux tests paramétriques ont une plus grande puissance de détection des effets que les tests non paramétriques correspondants. Ils permettraient de déceler un effet toxique dans un ensemble de données, même en présence d'irrégularités mineures, tandis qu'une analyse non paramétrique pourrait ne pas déceler l'effet.

78. Dans certains essais, on pourrait mesurer l'effet pour chacun des organismes (plusieurs) se trouvant dans une enceinte donnée (répétition). La comparaison intéressante serait celle des effets moyens à différentes concentrations. On estimerait à cette fin le rapport de : *a*) la variation à telles concentrations (c'est-à-dire entre les répétitions) ; *b*) la variation entre les concentrations. Les mesures se rapportant aux organismes individuels pourraient servir dans une analyse de variance si, pour quelque raison que ce soit, on voulait tester les différences entre les répétitions d'une même concentration, de même qu'entre les concentrations. Cela constituerait une analyse de variance « emboîtée », plus complexe, décrite dans des manuels de statistique.

Tableau P.2. — Présentation des résultats d'une analyse de variance hypothétique.

Source de variation	Somme des carrés des écarts à la moyenne	Degrés de liberté	Carrés moyens
Total	2 669	15 – 1 = 14	
Entre les concentrations	2 046	5 – 1 = 4	511,5
Entre les enceintes à la même concentration	623	5 (3 – 1) = 10	62,3

Pour ce concerne le tableau P.2, le véritable résultat d'une analyse de variance comporterait peut-être comme légendes des trois rangées les libellés « Total », « Inter » et « Intra » ou « Total », « Groupes » et « Erreur » plutôt que les légendes explicatives figurant dans le tableau. Dans la colonne des degrés de liberté, ne se trouveraient que les différences (14, 4, 10), sans explication arithmétique. On pourrait obtenir par soustraction les valeurs 623 et 10 de la 3e rangée.

Si le carré moyen « inter-concentrations » est plus grand que le carré moyen « intra-concentration », l'hypothèse nulle pourrait ne pas être vraie, c'est-à-dire qu'il y aurait une différence significative entre deux ou plusieurs traitements. On teste cela en divisant les carrés moyens « inter- » par les carrés moyens « intra- », le résultat étant désigné par *F*. Si *F* excède une valeur critique, fournie par le programme informatique ou trouvée dans les tables, il existe alors une différence significative quelque part entre les traitements (concentration).

Dans l'exemple hypothétique qui nous occupe, $F = 511,5/62,3 = 8,2$. La valeur critique de *F*, pour 4 et 10 degrés de liberté et $p = 0,05$, est 3,48. Comme la valeur calculée de *F* est supérieure à la valeur des tables, on rejette l'hypothèse nulle et on conclut qu'il existe une ou plusieurs différences entre les concentrations.

La comparaison de F *avec la valeur critique n'est valide que lorsqu'il a été satisfait aux hypothèses de l'analyse de variance.* Cela renvoie aux points soulevés dans le § P.2.4.

Si l'analyse de variance ne révèle aucune différence significative, elle se termine là, on accepte l'hypothèse nulle, et aucune toxicité n'a été prouvée. Si l'hypothèse nulle a été rejetée, il existe une différence, et l'analyse statistique passe au test de comparaisons multiples (§ 7.5 et P.4), afin de décider quels traitements différaient du témoin (et/ou de quels autres traitements).

En général, ces calculs sont effectués par un programme informatique tel que TOXSTAT, mais il est possible de les effectuer à la main, à l'aide des formules exposées dans Newman (1995) ou dans les manuels de statistique (Zar, 1974 ; 1999).

L'un des problèmes qui pourraient découler de l'analyse de variance serait d'avoir choisi une valeur erronée pour la somme des carrés des écarts à la moyenne de l'« erreur ». Si on avait effectué des mesures sur des organismes individuels à l'intérieur de la même répétition et si on avait saisi ces mesures dans l'analyse, le tableau P.2 comporterait des nombres supplémentaires, dans une autre rangée, ajoutée au bas du tableau. Sur les sorties d'ordinateur, cette rangée serait souvent légendée « Erreur ». L'expérimentateur pourrait, par mégarde, utiliser le carré moyen pour cette rangée dans le calcul de *F,* ce qui pourrait être correct dans certains autres plans d'expérience, comme il est mentionné dans la note de bas de page qui précède, mais qui n'est pas fréquent. Habituellement, on peut identifier assez facilement les bonnes valeurs dans le tableau imprimé et on peut les confirmer en examinant la ligne du tableau où le bon nombre de degrés de liberté est affiché.

Pour les tests d'hypothèses par analyse de variance, il est fortement souhaitable de posséder des échantillons de tailles égales (nombre égal de répétitions par traitement). En cas d'inégalité, l'analyse se complexifie, mais les programmes informatiques modernes s'en tirent bien et donnent la bonne valeur du terme de l'erreur pour tout test subséquent de comparaisons multiples. Dans le § 2.5, on mentionne d'autres aspects importants de la répétition. L'interprétation et les types d'erreur sont également pertinents (§ 7.2.2).

P.4 Tests paramétriques de comparaisons multiples

Nous avons décrit dans le § 7.5 l'emploi de tests de comparaisons multiples. Dans le § P.4, nous donnons des renseignements de base supplémentaires sur les tests. Il existe des instructions détaillées sur les marches à suivre pour les tests de comparaisons multiples (Hochberg et Tamhane, 1987).

P.4.1 Test de Williams

Le test de Williams est un test de comparaisons multiples recommandé pour servir principalement aux analyses paramétriques, après qu'une analyse de variance a montré l'existence d'une différence. Il possède une qualité importante, parce que, lorsque l'on compare chaque traitement au témoin, il tient compte de l'*ordre* des groupes selon la concentration croissante (ou décroissante) [Williams, 1972]. Cette information rend le test plus sensible. Le test de Williams est offert dans les programmes TOXCALC, TOXSTAT et CETIS.

Un exemple prouvera la sensibilité supérieure du test de Williams. Crane et Godolphin (2000) ont comparé les résultats précis d'essais du « laboratoire 1 » avec les résultats variables du « laboratoire 2 ». Il s'agissait d'observations hypothétiques de la mortalité avec trois répétitions, un témoin et 8 concentrations (exprimées en pourcentage d'effluent, c'est-à-dire 1,0, 2,2, 4,6, 10, 22, 46, 60 et 100 %). On a transformé les données en racines carrées et on les a analysées par analyse de variance et plusieurs tests de comparaisons multiples.

Les différences étaient frappantes. Non seulement les CSEO calculées différaient-elles étonnamment chez les deux laboratoires, mais, également, d'après les différents tests statistiques (tableau P.3). Le plus sensible des quatre tests a été celui de Williams (de 2 à 20 fois plus sensible que les autres). Il a été particulièrement efficace dans l'établissement d'une faible concentration pour les données variables du laboratoire 2.

Tableau P.3. — Différences dans les concentrations sans effet observé (CSEO) calculées au moyen de divers tests de comparaisons multiples. Les CSEO représentent le pourcentage d'effluent, pour les données hypothétiques, précises dans le cas du laboratoire 1 et variables dans le cas du laboratoire 2, présentées par Crane et Godolphin (2000).

Test de comparaisons multiples	CSEO (labo 1)	CSEO (labo 2)
Test de Williams	1,0	2,2
Test de Dunnett	2,2	22
Test *t* de Bonferroni	2,2	22
Test de Tukey	10	46

Le test de Williams opère par étapes. Il commence par la comparaison de l'effet de l'échantillon classé au premier rang (c'est-à-dire à la concentration maximale) avec l'effet observé chez les témoins, puis la comparaison de l'effet de l'échantillon du 2^{e} rang jusqu'à qu'aucune différence ne soit trouvée. Ainsi, il permet de trouver la plus faible concentration associée à un effet moyen significatif dans un groupe expérimental.

Le test de Williams est relié au test *t* et il partage les mêmes hypothèses. Les effets doivent être distribués d'une façon approximativement normale, les variances « intra-concentration » doivent être égales, et les observations doivent être indépendantes. Il aurait fallu satisfaire à ces exigences pour l'analyse antérieure de variance. Sinon, il conviendrait d'emprunter la voie non paramétrique, au moyen du test de Shirley (§ P.5.3) comme test correspondant à celui de Williams.

Le test doit s'appliquer à une suite monotone, c'est-à-dire que chaque effet moyen successif est soit : *a*) égal ou inférieur à l'effet précédent ; *b*) égal ou supérieur à l'effet précédent. Dans le cas où les suites ne seraient pas monotones, il existe une méthode de lissage qui pourrait devoir être appliqué à la main. Elle consiste à attribuer le même effet moyen aux deux effets moyens aberrants de la suite. La correction peut être appliquée plus d'une fois, si nécessaire, mais, dans la suite habituelle de résultats d'un essai de toxicité, cette « égalisation » des groupes pourrait faire perdre au test une partie importante de sa capacité de discrimination. L'expérimentateur s'apercevra facilement de ces situations lorsqu'il examinera les données originelles ou qu'il les traduira sous forme graphique ; dans ce cas, il devrait, pour débusquer les résultats anormaux, appliquer le test de Williams et, aussi, un autre test de comparaisons multiples.

Le test de Williams fonctionnera pour les nombres égaux ou inégaux d'observations contribuant à la valeur moyenne du témoin et de chaque traitement. Normalement, le terme de l'erreur calculé est obtenu à l'aide d'un programme informatique. Si un progiciel particulier ne peut pas fonctionner avec des nombres inégaux d'observations « inter-traitements », on peut effectuer les ajustements à la main. On a le choix entre deux formules simples pour les données équilibrées ou non équilibrées (Williams, 1972).

La valeur critique pour un ensemble particulier de données, correspondant aux degrés de liberté de l'erreur, peut être obtenue de tables fournies dans Williams (1971 ; 1972). Dans le cas des données non équilibrées, les valeurs critiques seraient obtenues des tables de Hochberg et Tamhane (1987). Dans la comparaison de la statistique calculée du test à la valeur critique, le premier à être inférieur de la valeur critique diffère significativement du témoin.

P.4.2 Test de Dunnett

Le test de Dunnett est un test usuel, par lequel on compare l'effet moyen de chaque traitement à l'effet moyen chez le témoin. Ce test jouit d'une certaine prééminence dans TOXSTAT, et dans les méthodes les plus courantes qui viennent des États-Unis [79]. Cependant, nous recommandons plutôt le test de Williams pour les essais d'Environnement Canada dont les résultats sont ordonnés (par ex. concentrations successives). Le test de Dunnett est moins puissant que celui de Williams pour la détermination de la CEMO parce qu'il ignore l'ordre des données (tableau P.3). En outre, dans la comparaison de tout traitement avec le témoin, il contrôle le taux d'erreur expérimentale plutôt que l'erreur se rapportant à une comparaison par paires.

Toutefois, le test de Dunnett est le choix qui convient pour une comparaison avec le témoin, quand il n'y a aucun ordre intrinsèque dans les traitements, c'est-à-dire que l'on ne s'attend à aucun gradient. Tel serait le cas, par ex., d'un essai sur un sédiment, si les matières provenaient d'un certain nombre d'emplacements différents, tous étudiés dans des échantillons répétés, mais seulement à une seule concentration, c'est-à-dire non dilués.

Le test de Dunnett exige que les données obéissent à la loi normale ; il représente une extension du test *t* (Dunnett, 1955 ; 1964). Il fait habituellement partie de progiciels visant l'exécution d'un *test unilatéral de signification,* ce qui répond à la situation prévue que les mesures correspondant aux concentrations expérimentales seront toutes dans le même sens par rapport à la mesure chez le témoin. Le test de Dunnett donne des résultats conservateurs (tendance à ne pas déceler de différences) pour les tests unilatéraux normaux.

---

79. Le logiciel du test de Dunnett est disponible à l'adresse http://www.epa.gov/nerleerd/stat2.htm.

Le test de Dunnett est habituellement appliqué aux expériences dont le nombre d'observations à chaque traitement est égal, et les vieux progiciels disponibles n'offrent que cette option. Parfois, les nombres d'observations pourraient être inégaux, par ex. plus d'observations chez le témoin. Le meilleur remède à cette situation serait de télécharger une version récente du test « modifié » de Dunnett (v. la note 79). Il existe aussi une modification appropriée, expliquée dans Newman (1995), et on trouve des exemples pratiques dans USEPA (1995). Les autres options, relativement aux nombres inégaux d'observations, sont les tests de Dunn-Sidak ou *t* corrigé par Bonferroni.

P.4.3 Ajustements de Dunn-Sidak et de Bonferroni pour des nombres inégaux de répétitions

Le test modifié de Dunnett est recommandé pour la comparaison de chaque traitement avec le témoin, quand le nombre d'observations est inégal. Si l'adaptation de ce test aux nombres inégaux d'observations n'était pas accessible, on pourrait se rabattre sur test de Dunn-Sidak. Nous mentionnons l'ajustement de Bonferroni parce qu'il est employé aux États-Unis, mais il ne confère aucun avantage particulier, et son utilisation n'a pas besoin d'être envisagée.

L'adaptation de Dunn-Sidak et celle de Bonferroni comparent la moyenne de chaque traitement à la moyenne du témoin. Aucune n'est très puissante par rapport au test de Williams, c'est-à-dire qu'elle ne pourrait ne pas permettre de distinguer des différences réelles. L'adaptation de Bonferroni est actuellement la norme dans les progiciels, tandis que celle de Dunn-Sidak est offerte dans le programme CETIS, TOXCALC et TOXSTAT, mais elle pourrait ne pas l'être dans certains progiciels. Un exemple pratique de l'adaptation de Bonferroni est offert dans USEPA (1995).

Les adaptations de Dunn-Sidak et de Bonferroni se fondent sur le test *t,* en apportant une correction aux valeurs critiques de *t*, pour tenir compte d'une comparaison multiple. Des comparaisons deux à deux répétées avec un test *t* normal pourraient aboutir à une erreur α (ou de première espèce) [§ 7.2.2]. Les progiciels effectuent automatiquement les corrections requises en effectuant une compensation quelque peu exagérée. La table des valeurs critiques que l'on peut utiliser pour le test de Dunn-Sidak peut être examinée, si on le désire, dans Newman (1995).

P.4.4 Tests de comparaison deux à deux

Des tests permettent de déceler la différence entre toutes les paires possibles de traitements. Bien que cette opération ne soit probablement pas nécessaire pour la plupart des essais de toxicité, cela pourrait être intéressant dans le cas d'essais sur le terrain ou d'une comparaison de divers emplacements. La *méthode LSD* (*Least Significant Difference*) *de Fisher* est apparentée au test *t* et est recommandée. Elle a l'avantage de contrôler l'erreur α se rapportant à une comparaison par paires, plutôt que l'erreur α expérimentale. La LSD peut servir pour les répétitions en nombre égal ou inégal. Elle n'est destinée qu'à un petit nombre de toutes les comparaisons possibles dans un ensemble de données, comparaisons qui seraient précisées d'avance et, à cet égard, semblables à d'autres tests de comparaisons multiples. La méthode LSD fait partie du progiciel SYSTAT et de quelques autres que l'on peut utiliser en toxicologie, et certains manuels la décrivent (Steel et Torrie, 1980 ; Steel *et al.*, 1997). Des instructions sur l'emploi du test sont données dans le § D.2.2 d'USEPA et USACE (1994).

En remplacement de la méthode LSD, on trouve, généralement disponibles dans les progiciels que l'on peut employer en toxicologie, le *test de Tukey* et *celui de Student-Newman-Keuls* (test SNK). Le test de Tukey peut fonctionner avec des échantillons de tailles inégales, bien que l'égalité soit souhaitable. Le test Tukey est peu sensible (tableau P.3).

P.5 Méthodes non paramétriques d'estimation de la CSEO

Si les résultats d'un essai ne peuvent pas satisfaire aux exigences de la normalité ni de l'homogénéité de la variance, même après transformation, on devrait les analyser par des méthodes non paramétriques, en employant les tests décrits dans le présent paragraphe et dans le § 7.5.2. Ces options non paramétriques sont de puissants outils à l'égard des données qui ne suivent pas la loi normale. Cependant, en général, ils seraient moins puissants pour la détection d'un effet toxique que les tests paramétriques correspondants, s'ils sont appliqués à des données obéissant à la loi normale.

Certaines méthodes non paramétriques exigent au moins quatre répétitions et parfois cinq[80]. C'est un fait reconnu dans les méthodes d'essais particuliers de toxicité sublétale publiées par Environnement Canada.

P.5.1 Tests initiaux d'hypothèse

Beaucoup de tests de comparaisons multiples non paramétriques sont « autosuffisants » et n'ont pas absolument besoin d'être précédés par un test qui serait analogue à l'analyse de variance. L'omission de cette étape initiale de test d'hypothèse a été courante en toxicologie. Cependant, nous recommandons de faire précéder beaucoup de tests de comparaisons multiples non paramétriques d'un test d'hypothèse(s) [v. la fig. 4]. Dans ces cas, l'analyse devrait passer à un test de comparaisons multiples, uniquement si le test initial rejette l'hypothèse de l'absence de différence entre les traitements. On veut, en effet, éviter de commettre des erreurs α dans la comparaison multiple. Autrement dit, le but est d'éviter de déclarer significative une différence entre deux traitements quand elle est le résultat du hasard, événement qui devrait survenir une fois sur 20 comparaisons de la p-valeur habituelle de 0,05. En termes de statistique, le test de comparaisons multiples est dit *protégé* par le test initial d'hypothèse qui opère un criblage. Ces tests en deux étapes constituent une approche prudente et, en principe, ils pourraient parfois aboutir à l'impossibilité de déceler une différence réelle (erreur de seconde espèce).

Dans les lignes qui suivent, nous décrivons trois de ces tests, à utiliser avec différents types de données non paramétriques (fig. 4). Ces tests sont les équivalents non paramétriques d'une analyse de variance (Zar, 1999) et ils montrent si, oui ou non, il existe au moins une différence entre les effets des traitements. Ces tests n'indiquent pas quel est l'effet différent des autres. Leur utilisation particulière dans différentes situations est montrée dans la fig. 4 et elle sera précisée dans les alinéas qui suivent.

Le test de la somme des rangs de Kruskal-Wallis (appelé ci-après test de Kruskal-Wallis) a été décrit par Kruskal et Wallis (1952). Il est parfois offert dans les progiciels (TOXSTAT, 1996) comme si c'était uniquement un test de comparaisons multiples, l'équivalent non paramétrique du test de Tukey. Cependant, ce test peut servir au test d'hypothèse(s) [analogue à l'analyse de variance] et aussi comme test de comparaisons multiples.

Le test de Fligner-Wolfe est un test de sommation des rangs que l'on peut utiliser pour tester une hypothèse nulle de l'absence d'effet (Fligner et Wolfe, 1982). Il vérifie l'hypothèse nulle selon laquelle aucune des médianes des traitements ne diffère de la médiane du témoin, l'hypothèse alternative étant que toutes les médianes des traitements sont plus grandes que la médiane du témoin. Cette hypothèse alternative diffère de l'hypothèse alternative habituelle avec de tels tests et elle est tout à fait explicite. Une conséquence sérieuse de cela est que le test ne convient pas lorsque certains traitements (concentrations) entraînent un effet mesuré supérieur et que certains entraînent un effet mesuré inférieur. Cependant, le test ne convient pas aux essais de toxicité hormétique, auquel cas on devrait utiliser le test de Kruskal-Wallis. L'autre limitation du test de Fligner-Wolfe est facile à surmonter. Si les traitements d'un essai de toxicité entraînent de plus faibles valeurs pour l'effet mesuré, on devrait multiplier toutes ces valeurs par – 1.

Le test de Jonckheere-Terpstra (Jonckheere, 1954) fonctionne également comme analogue non paramétrique de l'analyse de variance, et sa puissance est très grande. L'hypothèse nulle est que toutes les médianes sont égales et l'hypothèse alternative est un peu différente de l'hypothèse alternative habituelle, c'est-à-dire que les traitements sont ordonnés. En conséquence, le test convient très bien aux essais de toxicité. Bien qu'il soit offert dans certains logiciels de statistique importants, ce test, malheureusement, n'est pas encore offert dans les logiciels de toxicologie, et les calculs faits à la main sont très fastidieux et très longs.

---

80. La nécessité de disposer de quatre répétitions pourrait faire problème. Un plan d'expérience pourrait prévoir trois répétitions, principalement pour calculer une estimation ponctuelle, comme nous le recommandons dans ce document. Si l'expérimentateur voulait calculer la CSEO et la CEMO, cela pourrait se faire avec des méthodes paramétriques. Si, cependant, les résultats s'écartaient de la normalité et exigeaient une analyse par des méthodes non paramétriques, l'expérimentateur risquerait de ne pas pouvoir déterminer ces deux paramètres, selon le test non paramétrique particulier qu'il utiliserait. Dans ses méthodes récemment publiées, Environnement Canada exige quatre répétitions pour le test d'hypothèse(s), mais ce ne serait pas suffisant pour le test de Shirley.

P.5.2 Un test général de comparaisons multiples

Le test d'Edwards-Berry (Edwards et Berry, 1987) est un test de comparaisons multiples qui pourrait s'appliquer après n'importe lequel des trois tests que nous venons de mentionner pour le test d'hypothèse. Si l'hypothèse nulle était rejetée par suite d'un test quelconque, le test d'Edwards-Berry conviendrait à n'importe laquelle des situations décrites dans les alinéas qui suivent. Malheureusement, il n'est pas encore facile à trouver dans les progiciels, mais cela change. Le test d'Edwards-Berry utilise une technique de bootstrap pour créer une distribution empirique des données. Cela lui permet de manipuler la plupart des configurations de données, équilibrées ou non. Il produit une valeur critique qui « protège » le taux global d'erreur (lié à la famille de valeurs) [*family-wise comparison error rate*].

P.5.3 Données ordonnées — test de Shirley ou comparaison (deux à deux)

Le test de Shirley est une méthode non paramétrique très séduisante. Analogue au test paramétrique de Williams, il tient compte du classement des concentrations dans l'ordre croissant (ou décroissant). Il permet de comparer les effets à ceux que le témoin a subis et il n'est pas précédé par un test d'hypothèse (c'est-à-dire que l'on n'utilise pas d'analogue non paramétrique de l'analyse de variance ; v. la fig. 7. 1). Il est adaptable aux nombres inégaux de répétitions. Le test de Shirley est une extension du test de Kruskal-Wallis (v. § P.5.1), mais il devrait produire des résultats semblables à ceux du test de Williams. Le test pose par hypothèse que les effets décroissent monotonement et, sinon , on les lisse, comme dans le test de Williams. La taille de l'échantillon d'un traitement doit être d'au moins cinq.

Le test de Shirley classe les groupes selon le degré d'effet en utilisant les valeurs moyennes des effets chez le témoin et les groupes de traitements. Les valeurs réelles de la moyenne ne sont pas utilisées dans l'analyse comme elles le seraient dans le test de Williams. Le ou les effets observés chez les témoins sont classés dans la même suite que les traitements (concentrations expérimentales). Le test compare le rang moyen d'une concentration donnée au rang moyen du témoin. La variance est la variance non paramétrique des observations ordonnées. La méthode emploie la sommation des rangs. On compare le rang de la concentration maximale à celui du témoin. Si la comparaison fait conclure à une différence significative, elle passe à la concentration suivante (plus faible) tant qu'aucune différence n'est pas trouvée.

Le test de Shirley devrait être utilisé quand il sera accessible, mais, malheureusement, il ne fait pas partie de la plupart des progiciels employés en toxicologie et même de certains progiciels de statistique générale tels que SPSS (1996 ; 2001). La méthode n'est pas non plus décrite dans certains manuels usuels. Le test peut être exécuté à la main, bien que l'opération soit fastidieuse. Si le test n'est pas disponible, l'expérimentateur ayant besoin d'appliquer un test non paramétrique pourrait utiliser une comparaison deux à deux des données ordonnées (§ P.5.3) si les tests appropriés sont accessibles. L'autre possibilité, pour la comparaison avec le témoin seulement, serait d'utiliser les options pour un ensemble non ordonné de données, en commençant par le test de Fligner-Wolfe (§ P.5.4).

La comparaison par paires (deux à deux) de données ordonnées commence par un test d'hypothèse(s), employant le test de Jonckheere-Terpstra (§ P.5.1). Si on rejetait l'hypothèse nulle de l'absence de différence, l'analyse passerait ensuite au ***test de Hayter-Stone*** (Hayter et Stone, 1991). Ce test de contraste (test de comparaisons multiples) peut fonctionner avec des nombres égaux et inégaux de répétitions. Il existe des tables de valeurs critiques pour les petits et grands échantillons, si les répétitions sont en nombres égaux (c'est-à-dire des données équilibrées). Dans le cas des données non équilibrées, on dispose d'un nombre plus limité de valeurs critiques. Au moment d'écrire ces lignes, des tables de valeurs critiques n'existaient que pour les plus petits ensembles de données non équilibrées, y compris trois traitements ou moins et ne comptant pas plus de sept répétitions.

Les logiciels offrant le test de Jonckheere-Terpstra ou celui de Hayter-Stone ne sont pas faciles à trouver.

P.5.4 Comparaison, avec le témoin, de données non ordonnées

Dans le cas où les données ne sont pas ordonnées, nous recommandons le *test de Fligner-Wolfe* (§ P.5.1) pour vérifier l'hypothèse nulle de l'absence de différence d'avec le témoin. Si ce test n'est pas accessible dans un logiciel convenable,

on pourrait utiliser celui de Kruskal-Wallis. Si l'hypothèse nulle est rejetée et si les données sont équilibrées, le premier choix recommandé pour un test de comparaisons multiples avec le témoin multiple est le ***test de Nemenyi-Damico-Wolfe*** (Damico et Wolfe, 1987).

Le deuxième choix pour le test de comparaisons multiples est le ***test de sommation des rangs de Wilcoxon,*** généralement accessible, qui fonctionne avec un nombre inégal de répétitions. Le test de Wilcoxon résulte de la mise au point de méthodes et de valeurs critiques par un certain nombre de statisticiens (Newman, 1995).

Le test de Wilcoxon fonctionne semblablement au test multiunivoque de Steel (voir le texte qui suit). À une concentration donnée, on classe les différences entre les mesures expérimentales et les mesures chez les témoins correspondants. À chaque rang, on attribue un signe positif ou négatif, selon la nature de la différence par rapport au témoin. On somme les rangs positifs et, également, les rangs négatifs. On compare la plus petite des sommes positive et négative aux valeurs critiques connues pour déterminer s'il existe une différence significative entre l'effet expérimental et l'effet observé chez le témoin. La répétition de l'opération pour chaque concentration donne une estimation de la CSEO et de la CEMO. Ce test est généralement offert dans des logiciels. Un exemple pratique est donné dans USEPA (1995).

Un troisième choix est le ***test multiunivoque de Steel*** (Steel, 1959 ; 1961), offert dans la plupart des progiciels de statistique et diversement nommé. Un exemple pratique du test est donné dans USEPA (1995). La variante offerte dans les logiciels permet de ne manipuler que les données comportant un nombre égal d'observations à chaque traitement et chez le ou les témoins. Il faut au moins quatre observations (répétitions). Un progiciel offre un test unilatéral, c'est-à-dire que tous les échantillons renfermant le toxique sont réputés causer des effets identiques à ceux du témoin ou plus grands. Étant l'équivalent non paramétrique de test de Dunnett, celui de Steel peut servir à des comparaisons comme celles que nous avons mentionnées relativement aux essais sur un sédiment.

Le classement est au cœur de la méthode. On range (dans l'ordre croissant) huit mesures de moyennes : disons les quatre poids moyens correspondant à quatre répétitions, à une concentration donnée, avec les quatre mesures correspondant au témoin. On somme les rangs des mesures expérimentales ainsi que les rangs des mesures effectuées sur le témoin. On compare la plus petite des deux sommes des rangs à une valeur critique tirée d'une table usuelle. On déclare que, à cette concentration, les mesures expérimentales sont soit différentes, soit non différentes des mesures effectuées sur les témoins. On répète pour chaque concentration expérimentale cette énumération des valeurs en même temps que les valeurs mesurées chez le témoin. À la fin, l'expérimentateur sait quelles concentrations ont un effet significativement différent des effets observés chez le témoin (plus amples détails dans Newman, 1995). Il existe une modification pour le cas où toutes les concentrations expérimentales possèdent le même nombre d'observations, mais où le nombre d'observations correspondant au témoin est différent. Bien que cette modification ne soit pas disponible dans les progiciels usuels pour les essais d'écotoxicité, elle est décrite dans Newman (1995).

P.5.5 Comparaison par paires (deux à deux) de données non ordonnées

Le premier choix pour un test de comparaisons multiples est le ***test de Critchlow-Fligner-Steel-Dwass***, généralement appelé *test de Critchlow-Fligner* (Critchlow et Fligner, 1991) [81]. Ce test pourrait être utilisé si le test antérieur de Kruskal-Wallis avait mené au rejet de l'hypothèse selon laquelle tous les traitements ont révélé que les effets médians étaient égaux.

Le test consiste à comparer les résultats de chaque traitement à ceux que chaque autre traitement, y compris le témoin, et il révèle si les médianes sont égales ou différentes. Le test de Critchlow-Fligner pourrait être précédé du test de

---

81. Steel et Dwass ont proposé indépendamment un tel test de comparaison par paires, mais, dans chaque cas, il ne portait que sur des données équilibrées. Critchlow et Fligner (1991) ont élargi la porté du test aux résultats non équilibrés, de sorte que le nom des quatre est associé, comme il convient, à ce test.

Kruskal-Wallis (§ P.5.1) et il ne serait utilisé que si ce dernier menait au rejet de l'hypothèse nulle. Le test de Critchlow-Fligner convient aux nombres égaux ou inégaux de répétitions entre les traitements. C'est un test de comparaison bilatéral, c'est-à-dire qu'une différence pourrait être qu'un traitement présente des effets plus grands ou plus petits que ceux d'un autre traitement. Une comparaison donnée de deux traitements n'est pas influencée par les effets mesurés dans d'autres traitements ; cela est une caractéristique très séduisante dans un test non paramétrique de comparaisons multiples (Miller, 1981). Le test contrôle le taux d'erreur expérimentale, et il existe une faible probabilité de déclarer, à tort, qu'il existe une différence entre deux traitements.

Le test de Critchlow-Fligner n'est pas offert dans les progiciels usuels et il devrait être adapté à partir de sa description dans Critchlow et Fligner (1991). Les tables de valeurs critiques n'existent que pour un nombre limité de tailles d'échantillons, bien qu'il soit possible de produire les tables supplémentaires dont on a besoin.

Le test de comparaison par paires de Steel (Steel, 1960) est un second choix pour un test de comparaisons multiples et il convient aux données équilibrées. Si l'ensemble de données n'était pas équilibré et si le test de Critchlow-Fligner n'était pas disponible, le test de Kruskal-Wallis serait conscrit pour un double service. On l'utiliserait d'abord pour tester l'hypothèse nulle, puis, en cas de rejet, on l'utiliserait pour des comparaisons multiples afin de trouver quel ou quels effets correspondant à des traitements diffèrent de quels autres.

Différences statistiques entre les CE 50

On trouve dans le § 9.5.2 l'équation 9, qui pourrait servir de méthode pour effectuer le test du khi-deux sur les différences entre au moins trois CE 50.

$$\chi^2 = \sum_{i=1}^{k} w_i \left(\log(CE\ 50_i) - \left(\frac{\sum_{i=1}^{k} w_i \log(CE\ 50_i)}{\sum_{i=1}^{k} w_i} \right) \right)^2 \qquad (9)$$

On peut transformer l'équation 9 en l'équation Q.1, pour lui donner un aspect moins rébarbatif et montrer plus facilement les étapes des calculs.

$$\chi^2 = \sum_{i=1}^{k} w_i \left[\log(CE\ 50_i) - c \right]^2 \qquad (Q.1)$$

À son tour, *c* peut se définir par l'équation Q.2, et *b* peut se calculer de la façon montrée ci-dessous.

$$c = \frac{\sum_{i=1}^{k} b_i}{\sum_{i=1}^{k} w_i} \qquad (Q.2)$$

$$b = w \times \log(CE\ 50) \qquad w = \left(\frac{1}{s_{\bar{x}\,(\log CE\,50)}} \right)^2$$

Des exemples de calculs sont montrés dans le tableau Q.1. La comparaison se fonde sur trois des CE 50 qui ont servi d'exemples dans la note 63 du § 9.5.1. Les CE 50 et leurs limites de confiance sont : 8 (5,3 ; 12), 11 (7,3 ; 16,3) et 15 (10 ; 22,5), utilisés ici en tant qu'exemples A, B et C, respectivement. Pour les calculs du tableau Q.1, on remet les CE 50 sous leur forme logarithmique. Les erreurs types ($s_{\bar{x}}$) sont aussi calculées sur une échelle logarithmique et elles figurent dans la deuxième rangée du tableau. Dans les rangées inférieures, les calculs sont effectués conformément aux formules ci-dessus.

Il existe au moins une différence significative entre les trois CE 50, mais on ne sait pas où la ou les différences se trouvent. Un test approprié de comparaisons multiples n'a pas encore été conçu, mais il pourrait être mis au point (Zajdlik, en préparation). Encore, la méthode du tableau Q.1 pourrait-elle être utile. Elle précisera au moins les situations où une différence significative *n'existe pas*, évitant peut-être des questions inutiles sur la cause de différences qui, de fait, n'étaient pas significatives.

Tableau Q.1. — Exemples de calculs du khi-deux comme tests de l'existence de différences significatives entre trois CE 50.

	Essais de toxicité			Sigma (Σ) (= somme)
	A	B	C	
log (CE 50)	0,903 09	1,041 399	1,176 09	
Erreur type de log (CE 50) = $s_{\bar{x}}$	0,063 92	0,062 84	0,063 43	
$w = (1/s_{\bar{x}})$ élevé au carré	244,722 4	253,270 6	248,521 1	746,514 1
$b = w \times$ log (CE 50)	221,006 4	263,754 1	292,283 5	777,044 0
$c = \Sigma\, b / \Sigma\, w$				1,040 9
Log (CE 50) − c	− 0,137 81	0,000 50	0,135 19	
La valeur de la ligne ci-dessus élevée au carré	0,018 99	0,000 00	0,018 28	
La valeur de la ligne ci-dessus × w (= contributions au khi-deux)	4,647 44	0,000 06	4,542 37	9,190
Valeur critique de khi-deux pour 3 − 1 degrés de liberté et p de 0,05 = 5,991				
La valeur calculée (9,190) est supérieure à la valeur critique : il existe donc au moins une différence significative entre les trois CE 50.				

Une autre difficulté de l'emploi de cette méthode, actuellement, réside dans le fait que l'expérimentateur est rarement informé de la valeur de l'erreur type par les programmes statistiques usuels d'estimation de la CE 50. Les limites de la CE 50 sont toutefois fournies, et les manuels de statistique précisent la relation générale entre l'intervalle de confiance et l'erreur type (par ex. Zar, 1999). Cette relation est valide pour les types classiques de moyennes et de limites, et il n'est pas sûr que l'on peut l'utiliser pour les données sur la toxicité. Cette question pourrait être résolue par Zajdlik (en préparation).

Médiane et quartiles

La *médiane* et les *quartiles* sont des concepts mathématiques utiles et apparemment simples. Cependant, ces notions peuvent devenir déroutantes quand on tente d'appliquer les quartiles aux données de laboratoire. Nous expliquerons dans la présente annexe la nature de cette confusion, mais nous pourrions ne pas y remédier.

La médiane est uniformément définie, dans les publications, comme étant le nombre partageant en deux parties d'égale fréquence les éléments de la distribution supposée rangée par valeurs croissantes (ou décroissantes). Si la distribution compte un nombre impair d'éléments, la médiane est la valeur numérique du nombre ou de l'élément de rang $(n+1)/2$. Si le nombre d'éléments est pair, la médiane est la demi-somme des valeurs numériques des deux éléments de rang $n/2$ et $(n+1)/2$. Dans les deux cas, la médiane joue son rôle qui est de diviser la série, de sorte qu'*une moitié des éléments de la distribution rangée précède la médiane et l'autre moitié la suit.*

Les quartiles jouent un rôle semblable en divisant en plus petits groupes d'effectif égal une distribution ordonnée. Une règle empirique, l'*intervalle interquartile,* c'est-à-dire la différence numérique entre le *premier quartile* et le *3e quartile,* peut servir à reconnaître d'éventuelles observations aberrantes. Comme on le lit dans le glossaire, le quart des nombres d'une distribution ordonnés précéderaient le premier quartile, tandis que les trois quarts précéderaient le 3e quartile.

Nous recommandons à l'expérimentateur appelé à choisir les quartiles d'une distribution à retenir les valeurs les plus raisonnables pour satisfaire aux définitions générales exposées ci-dessus, c'est-à-dire les définitions d'un quart et de trois quarts. Parfois cela devient difficile à décider dans les distributions courtes et, dans ce cas, la définition de *quartile* devient moins utile, et on ferait mieux de l'éviter.

Quand une distribution compte peu d'éléments, le dilemme est aggravé par le fait que de nombreux mathématiciens réputés préconisent différents systèmes pour reconnaître ou calculer les quartiles. Huit variantes de méthodes ont été recensées (site Web www.xycoon.com/quartiles). On trouvera ci-dessous un exposé des méthodes usuelles, mais on peut voir les contradictions en cherchant *quartile* et *statistique* dans des sites Web apparemment universitaires.

La version le plus souvent trouvée de quartiles sera probablement celle qui se trouve dans les versions récentes des tableurs EXCEL et QUATTRO-PRO. Le premier quartile se calcule par la formule $L = 1/4\ (n + 3)$, tandis que le troisième quartile se calcule par la formule $U = 1/4\ (3n + 1)$ où n est le nombre d'éléments dans la distribution. Si le résultat du calcul est un nombre entier, ce nombre indique quel élément de la liste est le quartile. (Par exemple, si la réponse est 3, on choisit le 3e élément de la liste de nombres). Si le résultat est un nombre comportant une partie décimale, il désigne l'élément à choisir dans la liste et la proportion à calculer entre cet élément et l'élément suivant de la liste. (Par exemple, si la réponse est 3,75, on choisit le 3e élément de la liste et, par interpolation, on lui ajoute les trois quarts de la différence entre ce 3e élément et le 4e). EXCEL décrit la dernière situation comme suit : « Si le quartile se situe entre deux valeurs discrètes de la liste, la valeur fractionnaire est déterminée par l'interpolation linéaire ». On pourrait noter que s'il se trouve un nombre impair d'éléments dans la liste, cette méthode englobe la valeur médiane dans la moitié inférieure de la liste pour la détermination du quartile inférieur et, également dans la moitié supérieure pour le calcul du quartile supérieur. C'est en incluant ainsi la médiane que le statisticien John Tukey a défini les quartiles pour en permettre le calcul par des méthodes simples.

Cette méthode des tableurs peut être appliquée à la liste très courte (20, 24, 28, 34, 40) qui a servi d'exemple dans la note de bas de page du § 10.2. L'application de la formule de L donne 2,0 comme réponse, de sorte que le premier quartile est le 2e élément de la liste, c'est-à-dire 24. De même, la formule de U donne 4,0 comme réponse, de sorte que le quartile supérieur est le quatrième élément, c'est-à-dire 34. Un autre exemple serait une liste de huit éléments : 4, 5,

7, 9, 10, 12, 13, 16. L'application de la formule de *L* donne 2,75, de sorte que le premier quartile est 6,5. De même la formule de *U* donne 6,25, de sorte que le 3e quartile est 12,25. On conserve les décimales dans les valeurs des quartiles.

Statistique Canada recommande de choisir les quartiles de la même manière que pour la médiane, en précisant les formules à employer (site Web www.statcan.ca). Le premier quartile est l'élément qui partage en deux les observations ordonnées *inférieures* à la médiane. Le troisième quartile est l'élément qui partage en deux effectifs les observations ordonnées *supérieures* à la médiane. (La médiane ne fait partie d'aucune série.) Si on a un nombre pair d'éléments, on calcule le quartile selon la même méthode que celle qui a servi au calcul de la médiane. L'exemple donné est une série de 12 éléments, comme suit : 1, 11, 15, 19, 20, 24, 28, 34, 37, 47, 50, 57. La médiane est la demi-somme des 6e et 7e éléments, (24 + 28)/2 = 26. Sous cette valeur, on compte 6 éléments dont la valeur va de 1 à 24. Le premier quartile est la demi-somme des 3e et 4e valeurs, 15 et 19 = 17. Par un calcul analogue, le 3e quartile est 42. Statistique Canada ne donne aucun conseil sur la marche à suivre si le quartile est une valeur décimale, mais, vraisemblablement, cette valeur serait acceptée comme médiane.

Certaines sources sûres mentionnent que les quartiles doivent être des valeurs qui font effectivement partie de la série ordonnée de nombres. C'est cette méthode simple que nous avons suivie dans l'exemple de la note 64 du § 10.2. (Nous avons choisi les quartiles 24 et 34 dans la série 20, 24, 28, 34, 40.) L'utilisation de valeurs faisant effectivement partie de la série résulte des formules décrites par Zar (1974), qui sont semblables aux formules des tableurs auxquelles nous avons fait allusion. L'expression (n + 1)/4 identifie l'élément qui est le premier quartile. De même 3 (n + 1)/4, ou 0,75 (n + 1), identifie l'élément qui est le 3e quartile. Si le résultat n'est pas un nombre entier (par ex. 1,5), on choisit l'entier supérieur immédiat (par ex. le 2e élément de la série). Cette sélection peut faire problème si la série compte peu d'éléments, comme celle que nous avons énumérée ci-dessus (20, 24, 28, 34, 40) ; le premier quartile est le 2e élément de la série (24), mais le 3e quartile serait le 5e élément de la série, c'est-à-dire 40. Ce 5e et dernier élément de la série joue à peine le rôle qui est le sien, c'est-à-dire d'avoir le quart des valeurs au-dessus de lui.

Dans certaines des premières versions du tableur EXCEL de Microsoft, on a utilisé une variante de la méthode qui précède. Si le calcul donnait une valeur décimale dont les décimales seraient 51 ou plus, on choisirait l'élément supérieur suivant de la série (par ex. pour 1,51, ce serait le 2e élément). Si les décimales calculées étaient 49 ou moins, on choisirait l'élément dont le rang correspondrait à la partie entière (par ex. pour 1,49, ce serait le 1er élément). Si le calcul avait donné une valeur décimale dont les décimales seraient 50, on choisirait la valeur calculée comme quartile (par ex. 1,5). Si le calcul donnait un entier comme résultat, on choisirait l'élément dont le rang égale l'entier en question (par ex. pour 2,0, ce serait le 2e élément).

Parfois, on trouve le conseil suivant : omettre le terme « + 1 » des formules décrites par Zar (1974) et données ci-dessus. Le premier quartile répondrait à la formule n/4 et le troisième par 3n/4, en arrondissant à l'entier supérieur suivant au besoin. Pour la série de cinq éléments employée dans l'exemple ci-dessus, les quartiles seraient le 2e et le 4e éléments, c'est-à-dire 24 et 34 comme dans la note 64, § 10.2.

Certains augmentent l'exactitude de la sélection en autorisant des valeurs qui auraient pu se trouver dans la série. Par exemple, dans une série d'entiers, un quartile serait un entier, mais il pourrait, de fait, ne pas se trouver dans la série.

Devant tous ces conseils divergents, nous recommandons, s'il faut choisir un 1er et un 3e quartiles, de choisir les valeurs les plus sensées, qui répondent à la définition du glossaire, c'est-à-dire diviser la série de nombres en quatre parties égales. Dans le cas des grandes séries de nombres, les conseils contradictoire ne feront pas problème : toutes les méthodes donneront des réponses semblables et la méthode du tableur sera satisfaisante. Dans le cas des séries courtes, nous ne recommandons pas l'emploi de quartiles.